Technische
Universität
Braunschweig

Data-driven variable-fidelity reduced order modeling for efficient vehicle shape optimization

Von der
Carl-Friedrich-Gauß-Fakultät
der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades einer
Doktoringenieurin (Dr.-Ing.)
genehmigte

Dissertation

von
Anna Bertram geb. Sauerbrei
geboren am 03.07.1989
in Gifhorn.

Eingereicht am:	04.06.2018
Disputation am:	21.09.2018

1. Referent:	Prof. Dr. Ralf Zimmermann
2. Referent:	Prof. Dr. Matthias Bollhöfer

2018

Bertram, Anna
Data-driven variable-fidelity reduced order modeling for efficient vehicle shape optimization
Dissertation,
Technische Universität Carolo-Wilhelmina zu Braunschweig, 2018.

The results, opinions and conclusions of this thesis are not necessarily those of the Volkswagen AG.

Zusammenfassung

In den letzten Jahrzehnten hat die Bedeutung der Aerodynamik in der Fahrzeugentwicklung immer weiter zugenommen. Um Kraftstoffverbrauch, Emissionen und die Umweltbelastung zu reduzieren, ist die Automobilindustrie bestrebt die Effizienz von Fahrzeugen zu verbessern. Ein Aspekt, um dieses Ziel zu erreichen, ist die Optimierung des Luftwiderstands neuer Fahrzeuggeometrien. Zur Untersuchung des aerodynamischen Verhaltens neuer Geometrien kommen Computersimulationen zum Einsatz. Typische Zielgrößen sind dabei skalare Größen, wie der Widerstandsbeiwert, aber auch hoch-dimensionale Feldgrößen wie die Druckverteilung auf der Fahrzeugoberfläche, das Luftgeschwindigkeitsfeld über dem Fahrzeug oder auch die sogenannte Sensitivitätslandkarte, die die Sensitivität des Widerstandsbeiwertes bezüglich lokaler Formänderungen anzeigt. Allerdings benötigt das Erstellen hochgenauer aerodynamischer Daten mittels Strömungssimulationen – selbst unter Verwendung von modernen Hochleistungsrechnern – heutzutage circa eine Woche, was die Verwendung dieser Simulationen in einem Optimierungsprozess nahezu unmöglich macht. Aus diesem Grund werden Ersatzmodelle benötigt, die die gewünschte Strömungsgröße mit weniger Rechenaufwand näherungsweise berechnen. Da das Design von Autos weiterhin in erster Linie durch ästhetische Gesichtspunkte bestimmt wird, ist außerdem eine automatisierte Optimierung der Fahrzeugform von der Industrie nicht gewünscht. Zum Einsatz in der aerodynamischen Formoptimierung bedarf es daher eines interaktiven Design-Tools, welches aerodynamische Zielgrößen hinreichend genau und möglichst in Echtzeit näherungsweise berechnet und visualisiert.

Um ein Ersatzmodell zu entwickeln, welches diesen Anforderungen genügt, kommen wegen ihrer schnellen Rechenzeit nicht-intrusive Modellreduktions-Verfahren infrage. Die Strömungssimulation wird dabei als Black-Box-Modell betrachtet, welches eine Menge von Eingangsparameter auf die Zielgröße abbildet. Die gewünschte aerodynamische Zielgröße für eine vorgegebene Fahrzeuggeometrie wird dann auf Grundlage von beobachteten Stützwerten der Zielgröße approximiert, welche zuvor mittels Strömungssimulation für ausgewählte Eingangsparameter berechnet wurden. Abhängig von der Komplexität des Zusammenhangs zwischen Eingangsparametern und Zielgröße sowie der Anzahl der Eingangsparameter bedarf es zur Erstellung eines hinreichend genauen Ersatzmodells ausreichend vieler dieser Stützwerte. Aufgrund der hohen Rechenzeit hochgenauer Strömungssimulationen ist die Anzahl der verfügbaren Stützwerte in der aerodynamischen Formoptimierung allerdings stark limitiert.

Im Falle skalarer Zielgrößen kann man in solchen Situationen mithilfe von sogenannten *variable-fidelity* Verfahren die Effizienz der Erstellung von Ersatzmodellen drastisch steigern. Neben dem rechenintensiven, hochgenauen Black-Box-Modell basieren diese Techniken zusätzlich auf einem zweiten Computermodell zur Berechnung der Zielgröße, welches sich schneller auswerten lässt, aber ungenauere Rückgabewerte liefert. Solche Computermodelle sind häufig durch physikalische oder numerische Vereinfachungen des hochgenauen Modells verfügbar. Stützwerte des weniger genauen Modells werden dann genutzt um den globalen Verlauf des Zusammenhangs zwischen Eingangsparametern und Zielgröße zu erfassen und dadurch die Approximation der Zielgröße basierend auf den hochgenauen Stützwerten zu verbessern.

Im Rahmen dieser Arbeit wird eine bekannte variable-fidelity Methodik für skalare Zielgrößen auf vektorwertigen Zielgrößen erweitert. Es wird gezeigt, dass der vorgestellte neue Ansatz eine Verallgemeinerung bekannter variable-fidelity Verfahren für skalare Zielgrößen ist und sich dadurch bestimmte Eigenschaften dieser bekannten Ansätze auf die neue Methode übertragen. Ein Fehlerschätzer wird hergeleitet, der zur adaptiven Verbesserung des Modells verwendet werden kann. Außerdem werden offene theoretische Fragen zu der Cokriging-Methode, einer bekannten variable-fidelity Methode für skalare Zielgrößen, untersucht. Anschließend wird die Anwendbarkeit der neuen Methode in der industriellen Fahrzeugform-Optimierung in zwei Fallstudien am Beispiel eines hochaufgelösten Computermodells eines Volkswagen Passat B6, welches von der Volkswagen AG zur Verfügung gestellt wurde, demonstriert und mit herkömmlichen Verfahren verglichen.

Danksagung

Es gibt viele Menschen, ohne die die Fertigstellung dieser Arbeit so nicht denkbar gewesen wäre und die daher an dieser Stelle nicht unerwähnt bleiben sollen.

An erster Stelle bedanke ich mich bei Prof. Dr. Ralf Zimmermann für all seine Unterstützung und seine Zeit. Vielen Dank für die vielen Hinweise und Diskussionen, jedes offene Ohr sowie den ein oder anderen Rat abseits der Mathematik.

Eine wichtige Rolle spielt natürlich auch das richtige Arbeitsumfeld, daher gilt besonderer Dank meiner Arbeitsgruppe, der AG Numerik. Allen voran danke ich Prof. Dr. Heike Faßbender, die mich immer unterstützt und mir vieles ermöglicht hat. Ich danke Prof. Dr. Matthias Bollhöfer für sein Interesse und so manche fachliche Debatte. Ich danke Tanja für ihre Organisation und ihren Beistand in schwierigen Phasen. Meiner Büro-Nachbarin Lena gebührt mein Dank fürs Zuhören und den ein oder anderen Tipp aus einer anderen Blickrichtung. Bedanken möchte ich mich weiter bei Nikta für ihre Stärke und für das Händchenhalten vor meinem ersten Vortrag auf einer internationalen Konferenz. Ich danke Philip für jedes Matrixpolynom sowie das gemeinsame Aushecken von wunderschönen Theorien und ganz besonders Christian für das Zerstören eben dieser und die Versuche die Sache doch noch irgendwie zu retten. Ich danke euch allen für jede gemeinsame Stunde in und außerhalb des Büros. Es war eine wunderschöne Zeit.

Die Mathematik in die Anwendung zu bringen reizt mich seit meiner Schulzeit und treibt mich an. Aus diesem Grund möchte ich mich ganz besonders bei meinen Partnern von der Volkswagen AG für die große Unterstützung, die vielen Diskussionen, die Begeisterung, die Zeit und überhaupt die Möglichkeit zur industrienahen Forschung in diesem Promotionsprojekt bedanken.

Ich danke einmal mehr Prof. Dr. Ralf Zimmermann sowie Prof. Dr. Matthias Bollhöfer für die Übernahme und das schnelle Erstellen der Referate zu dieser Arbeit und Prof. Dr. Jens-Peter Kreiß für die Übernahme des Vorsitz des Promotionsausschusses sowie die flexibele und unkomplizierte Terminabsprache.

Da es bei einer Promotion wie im gesamten Leben nicht immer nur bergauf geht, braucht es Familie und Freunde, die an einen glauben. Ein besonderer Dank gilt daher meinen Freunden für den vielen Zuspruch, die Ablenkung und den nötigen Ausgleich zur Mathematik. Ich danke meiner Familie, die an mich geglaubt hat, mir die nötigen Freiheiten einräumte, mich immer wieder aufgebaut hat und mir den Rückhalt gab, um diese Arbeit erfolgreich fertig zu stellen. Ohne euch hätte ich das nicht geschafft.

Contents

List of Figures

List of Tables

List of Algorithms

Nomenclature

Abbreviations

BLUE	best linear unbiased estimator
cf.	compare (from lt. *confer*)
CFD	computational fluid dynamics
DES	detached-eddy simulation
DNS	direct numerical simulation
e.g.	for example (from lt. *exempli gratia*)
et seq.	and the following (from lt. *et sequentes*)
i.a.	among others (from lt. *inter alia*)
i.e.	that is to say (from lt. *id est*)
IAD	interactive aerodynamic design
LES	large-eddy simulation
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
MOR	model order reduction
PDE	partial differential equation
POD	proper orthogonal decomposition
RANS	Raynolds-averaged Navier-Stokes
resp.	respectively
RSM	response surface model
s.p.d.	symmetric positive definite
SVD	singular value decomposition
w.l.o.g.	without loss of generality

Mathematical symbols and notation

$\mathbb{C}$	set of complex numbers
$\mathcal{D}$	design space, $\mathcal{D} \subset \mathbb{R}^d$
$\mathbb{N}$	set of natural number
$\mathbb{R}$	set of real numbers

1. Introduction

Over the last decades, aerodynamics plays a more and more important role in engineering design of vehicles. In order to reduce fuel consumption, emissions and the burden on nature, the automotive industry is making efforts to improve the efficiency of vehicles. One aspect to achieve this goal is the optimization of the aerodynamic drag of new vehicle geometries. To study the aerodynamic behavior of new designs, computer simulations are frequently used. Typical quantities of interest in such studies are the scalar-valued drag coefficient, which is often used to quantify the aerodynamic drag of a body, and field variables like the pressure distribution on the surface of the car or the velocity field above the car. Of particular interest is also the so-called *sensitivity map*, a vector-valued quantity which indicates the sensitivity of the drag coefficient with respect to local shape variations. There are two main issues related to car shape optimization, [5]:

(1) From a mathematical point of view, the flow around a body can be described via the *Navier-Stokes Equations*, a system of nonlinear coupled partial differential equations (PDEs). This system of equations is extremely hard to solve numerically. Different simplifications can be applied which lead to a variety of numerical methods of different complexity and accuracy. The computational effort to obtain sufficiently accurate results remains, however, huge: Nowadays, even if high-performance computers are used, the generation of accurate aerodynamic data (including model preparation, computation and post-processing) requires of the order of a full week.

(2) The design of a car is primarily driven by aesthetic considerations rather than aerodynamic optimality. Since aesthetic ideas on the design of a car can hardly be expressed in mathematical equations, an automatic optimization of the shape is not an option. Instead, one is interested in an *Interactive Aerodynamic Design* (IAD) tool, which enables the computation of car aerodynamics in near real-time.

This calls for the development of a so-called *surrogate model*, which yields sufficiently accurate approximations of the quantity of interest and, simultaneously, features real-time-capable evaluations.

Surrogate models are educated guesses of how a computational mapping between input and output data might look like. They can be used whenever a computer model for a desired quantity is not available or an evaluation is too expensive for optimization purposes.

In literature it is distinguished between three different classes of surrogate models, [2, § 1.2, p. 486 et seq.]: hierarchical models, projection-based reduced order models and data-fit models. Hierarchical models are surrogate models which are driven by physics. They result from the original model, for instance, by considering simplifying assumptions on the physics (e.g. neglecting friction), a coarser discretization of the spatial domain or relaxed convergence criteria for an iterative solver. Models of the second class result from the projection of the governing equations onto a lower-dimensional subspace. This has the advantage that the structure and thus parts of the physical meaning of the underlying model are retained. On the other hand, projection-based reduced order models are intrusive: Their construction requires access to the underlying computer code of the original model. However, especially in industrial application fields, the original computer model is often a software bought by the company and the code is not directly accessible. Therefore, projection-based reduced order models can often not be easily implemented in a running industrial process chain.

In contrast to hierarchical and projection-based surrogate models, data-fit models use regression or interpolation techniques to fit a model based on sampled data. Data-fit modeling techniques are fully data-driven and treat the underlying original model as a black box, which maps input variables to output variables. Regarding industrial applications, this non-intrusiveness is a significant benefit over projection-based surrogate models since no access to the source code is required. Furthermore, due to their independence from the original computationally expensive model, non-intrusive surrogate models feature remarkable fast evaluation times. Certainly, the quality and thus the usability of a data-fit surrogate model depends on the information and data which is available to it: Assumptions on the smoothness, the global trend, the number and the location of sampled data, gradient information, etc. Especially if the evaluation of the original model is computationally expensive—as is the case with aerodynamic simulations—, available data is limited. This issue can be addressed by combining data-fit models of different levels of accuracy. Since hierarchical surrogate models with low computational costs for evaluation are often available for a certain quantity, one can afford more sampled data of these “low-fidelity” models. Even though the quality of this sampled data is not good enough to yield a sufficiently accurate data-fit model, it can be used in addition to sampled data of the original model to obtain information about the global trend of the quantity of interest. This idea forms the basis of the so-called *variable-fidelity* surrogate models.

1.1. State-of-the-art

The state-of-the-art method in surrogate modeling of nonlinear large-scale dynamical systems is proper orthogonal decomposition (POD), also known as principal component analysis or Karhunen–Loève decomposition, [63]. It is strongly related to the singular value decomposition of a real matrix. Based on sampled data of the original model, POD seeks for a subspace of a pre-chosen dimension, which best approximates the sampled data[1]. The computation of the quantity of interest is then restricted to this subspace. Surrogate modeling techniques which use POD can be divided into three categories depending on the way a solution in the low dimensional subspace is obtained, [30]: Galerkin projection, residual minimization and interpolation of the basis coefficients. In standard POD-Galerkin methods, the governing equations of the original model are projected onto the lower dimensional subspace resulting from the POD. This procedure has been successfully applied to the prediction of aerodynamic data in the last decades, [4, 69, 75]. On the other hand, residual minimization based methods seek for an approximation to the quantity of interest by solving a low-dimensional optimization problem in the POD subspace, [50, 83].

However, if the underlying system features strong nonlinearities, the computational effort associated with these two approaches remains high. This problem is addressed by two state-of-the-art modifications to the original approaches: The Gauß-Newton with approximate tensors (GNAT) method, [12, 13], and the discrete empirical interpolation method (DEIM), [15]. The GNAT method uses the Gauß-Newton method and a Petrov-Galerkin projection in a residual minimization based approach to reduce the dimensionality of the nonlinear terms. On the other hand, DEIM is a modification of the Galerkin-POD approach and uses an additional POD subspace to reduce the nonlinearities.

POD-Galerkin and residual minimization based approaches belong to the class of projection-based reduced order models, which have the drawback of being intrusive. On the other hand, POD methods which are based on the interpolation of the basis coefficients are fully data-driven and do not require access to the computer code of the original model. The idea is to express the sampled data by a linear combination of the POD basis vectors and to interpolate the basis coefficients. The latter is realized by using a data-fit surrogate modeling approach for scalar-valued responses, [9, 54, 78, 80], which are reviewed below. This approach was first proposed

---

[1] Instead of choosing the dimension of the subspace in advance, a common approach is to determine it so that a preset amount of *information* given by the sampled data is retained, [63, Sec. 2.2].

in [54] for steady-state quantities and later extended to the unsteady case in [78]. An important observation was made in [80], for the case that Kriging or radial basis function interpolation is used to obtain surrogate models for the basis coefficients: Under certain assumptions the interpolation of the POD basis coefficients yields exactly the same surrogate model as entry-by-entry interpolation of the vector-valued sampled data.

For the data-fit surrogate modeling of scalar-valued responses, sometimes also referred to as *response surface modeling* (RSM) in literature, [67], a variety of approaches exists among which the most popular are polynomial and piecewise polynomial regression and interpolation models, [28, 59], radial basis functions, [10, 14], and the statistical interpolation method Kriging. The latter is also well known as the spatial linear model or Gaussian process regression.

The Kriging method is based on the assumption that the original model is a realization of a random process. A linear unbiased estimator is constructed such that the mean squared error is minimized. Kriging has its origin in the geostatistic community, [40, 46, 55], but gained popularity over the last decades because of the increasing importance of computer experiments. Due to its capability to handle highly nonlinear responses and its fast evaluation times, it is nowadays extensively used in many different areas. Among others it is applied in the design and analysis of computer experiments [44, 53, 70, 71], machine learning [65] and surrogate modeling [28]. In the latter context it has been successfully applied to different application fields, e.g. geographical information systems, [60], modeling of ozone levels, [1], and aerodynamic applications, [25, 39, 48].

In the context of scalar-valued responses, the simplest way to incorporate data from different levels of accuracy is by introducing a so-called bridge function, to correct a surrogate model for the low-fidelity data to approximate the high-fidelity model, [35]. Depending on the type of correction, one distinguishes between additive, multiplicative and hybrid bridge function methods.

One method which is able to handle more complex relationships than the simple bridge function methods, is called *Cokriging*, [28, p. 177]. Cokriging is a direct extension of Kriging to the variable-fidelity framework and was—as Kriging—originally developed by the geostatistic community, [40]. As with Kriging, it is assumed that the original high- and the less accurate low-fidelity model are realizations of correlated random processes. However, the original formulation of the Cokriging method gained hardly any attention in the surrogate modeling of computationally expensive computer outputs. This can be explained by difficulties in predicting the cross-covariances between the high- and low-fidelity outputs. A new Bayesian approach was introduced in [42] to address this issue. The authors use an autoregressive model to determine the cross-covariances. The approach was successfully applied in the context of aerodynamic optimization in [25–28]. A new Cokriging formulation which follows closely the common formulation of Kriging in surrogate modeling was introduced and tested for the prediction of aerodynamic data in [36, 37]. In order to introduce a robust alternative to Cokriging which is easier to implement based on an existing Kriging code, a method called *Hierarchical Kriging* was developed and tested in aerodynamic applications by the same group, [34].

Another Kriging-based surrogate modeling technique, which allows to include gradient information, is known as *Gradient Enhanced Kriging*, [58]. A recent overview of publications on Gradient Enhanced Kriging as well as a comparative study with other gradient enhanced surrogate modeling techniques is given in [49]. To enable the processing of gradient information in the surrogate modeling of vector-valued quantities, a method which combines Gradient Enhanced Kriging with POD was introduced in [79].

While there exist numerous variable-fidelity approaches for the surrogate modeling of scalar-valued quantities like Cokriging and Hierarchical Kriging, there is, apart from the preliminary work included in the thesis [56], no such approach for the modeling of vector-valued quantities.[2]

1.2. Goal of this work

For the use in aerodynamic vehicle shape optimization, we seek for a surrogate modeling technique which enables the approximation of high-dimensional aerodynamic quantities at a sufficient level of accuracy in near real-time. The requirements of fast evaluation times and integrability in industrial process chains calls for a non-intrusive approach as given by methods based on the interpolation of the POD basis coefficients. However, the tremendous computational cost associated with a high-fidelity aerodynamic simulation heavily limits the number of available high-fidelity sample points. Thus, the application of existing methods like POD combined with Kriging basis coefficient interpolation is strongly restricted to cases with weak non-

---

[2] We published a first case study on variable-fidelity surrogate models for high-dimensional quantities in January 2018, [5]. Independent of our work, a related variable-fidelity approach for high-dimensional outputs was developed and published in [66] after this thesis was submitted.

linearities or a small number of input variables. To address this issue, a new surrogate modeling technique is needed which features accurate surrogate models at reasonable computational effort. To this end, we enlarged the framework of POD basis coefficient interpolation to handle data of different levels of accuracy. Variable-fidelity interpolation techniques like Cokriging or Hierarchical Kriging are then used to interpolate the basis coefficients. We investigate the new approach in terms of two realistic industrial case studies and compare the performance with the single-fidelity modeling approach. Furthermore, theoretical questions associated with this procedure are addressed:

(1) For the Kriging and radial basis function interpolation of the POD basis coefficients it is known that under certain assumptions, exactly the same surrogate model is derived as via entry-by-entry interpolation of the vector-valued sampled data. What can be said about the new variable-fidelity surrogate modeling approach?

(2) Is there a reliable error estimate which enables an adaptive improvement of the surrogate model by iteratively adding new sample points?

1.3. Overview

The thesis is structured as follows. In Section 2, the mathematical basics of the surrogate modeling techniques considered in this work are discussed. Some basics of statistics will be needed and are therefore revised in Section 2.1, followed by an introduction to the theory of positive definite functions in Section 2.2. A short discussion on the Schur complement and conditions for positive definite matrices are given in Section 2.3. As stated above, the basic concept of the proper orthogonal decomposition is strongly related to the singular value decomposition of a real matrix. Thus, an overview of the singular value decomposition is provided in Section 2.4.

Due to the aim of approximating aerodynamic quantities, Section 3 deals with basics of fluid dynamics. The governing equations of fluid dynamics, the Navier-Stokes Equations, are explained in Section 3.1 by applying the fundamental laws of continuum mechanics to a fluid element: the conservation of mass, momentum and energy. The Navier-Stokes equations are extremely hard to solve numerically. A direct numerical simulation is often not feasible for realistic flow problems. Alternative numerical simulation methods which are currently state-of-the-art in industry are the Reynolds-averaged Navier-Stokes simulation, the large-eddy simulation and the detached-eddy simulation. These methods are described in Section 3.2. A quantity which is commonly used to quantify the aerodynamic drag of a body is the drag coefficient. A definition is given in Section 3.3.

In order to give an overview, all steps of constructing a data-fit surrogate model are discussed in Section 4.

Interpolation techniques for scalar-valued quantities are in the focus of Section 5. Here, the Kriging method is presented in detail in Section 5.1. Cokriging is a direct extension of Kriging to the variable-fidelity framework. After the method is introduced in Section 5.2.1–5.2.2, open theoretical questions are addressed in Section 5.2.3 and Section 5.2.4. The findings are validated by means of numerical experiments in Section 5.2.5. Finally, a model parameter is investigated numerically in Section 5.2.6. Another variable-fidelity method based on Kriging, the Hierarchical Kriging method, is presented in Section 5.3.

In Section 6, the variable-fidelity methodology is extended to the case of vector-valued quantities of interest. A brief introduction to proper orthogonal decomposition is given in Section 6.1. The new variable-fidelity approach is described in Section 6.2, followed by some theoretical remarks on this approach in Section 6.3. Error estimation and adaptive sampling strategies are covered in Section 6.4.

The introduced method is investigated by means of numerical experiments in Section 7. The general applicability of the variable-fidelity modeling approach and the adaptive sampling strategy is demonstrated by means of a benchmark problem in Section 7.1. To further investigate the method with regard to the usability in industrial vehicle shape optimization, two case studies were carried out based on a realistic high-resolution computer model of a Volkswagen Passat B6[3]. In the first case study in Section 7.2, the goal is to obtain a surrogate model for the pressure distribution on the surface of the vehicle as a function of 5 design parameters introduced at the rear roof. High- and low-fidelity sampled data is computed via Reynolds-averaged Navier-Stokes simulations of different stages of convergence. Contrary, in the second case study in Section 7.3, detached-eddy simulations were carried out to obtain high-fidelity data, while Reynolds-averaged Navier-Stokes simulations were used to obtain low-fidelity data. Again, the pressure distribution on the surface of the vehicle serves as quantity of interest. In order to increase the complexity of this test case, three design parameters at the rear of the vehicle were allowed to vary in an extreme and non-realistic range.

The thesis is concluded in Section 8 and an outlook on further work is given.

Additional information on the implemented and utilized software can be found

---

[3] kindly provided by Volkswagen

in Appendix A.

2. Mathematical background

Before we start with the main part of this thesis, some mathematical basics are recapitulated in this Section, which are needed for the concepts and proofs in the following parts. For surrogate modeling we will in particular consider statistical interpolation methods, meaning that the spatial correlation of the input data is taken into account. Beside some basics of statistics repeated in Sec. 2.1, *positive definite functions* are needed, as they play an import role in the modeling of these spatial correlations. They are discussed in Sec. 2.2. The term of positive definite functions is, as the name suggests, closely related to the well-known concept of positive definiteness of matrices. To determine whether a block matrix has this property, the *Schur complement* is an useful tool and is therefore addressed in Sec. 2.3. The Section is closed by some basics on the *singular value decomposition* in Sec. 2.4, which is due to its properties of particular importance in model order reduction.

2.1. Basics of statistics

As stated above, some basics of statistics are needed and therefore repeated hereafter. We start with the definition of a few fundamental terms for real-valued random variables. They are taken from [45, p. 52 et seqq., p. 145].

Definition 2.1. *Let $X\colon \Omega \to \mathbb{R}$ be a real-valued random variable and suppose that the expected values of X and X^2 exist. The* variance *of X is defined as*

$$\operatorname{Var}[x] := \operatorname{E}\left[(X - \operatorname{E}[X])^2\right] \tag{2.1}$$

and the quantity

$$\sigma_X := \sqrt{\operatorname{Var}[X]} \tag{2.2}$$

is called standard deviation *of X. If $Y\colon \Omega \to \mathbb{R}$ is a second real-valued random variable with existing expected values for Y and Y^2, then the* covariance *of X and Y is*

$$\operatorname{Cov}[X,Y] := \operatorname{E}\left[(X - \operatorname{E}[X])(Y - \operatorname{E}[Y])\right] \tag{2.3}$$

and the correlation *of X and Y is defined as*

$$\operatorname{Cor}[X,Y] := \frac{\operatorname{Cov}[X,Y]}{\sigma_X \sigma_Y}. \tag{2.4}$$

The concept of expected value and variance of a random variable can be extended to vectors of random variables as follows:

Definition 2.2. *Let* $Z = (Z_1, \dots, Z_n)^{\mathrm{T}}$ *be a vector of real-valued random variables and suppose that for every random variable* Z_i*,* $i = 1, \dots, n$ *the expected values* $\mathrm{E}[Z_i]$ *and* $\mathrm{E}[Z_i^2]$ *exist. The* expected value *of the vector* Z *denotes the vector of expected values of the random variables* Z_i*,*

$$\mathrm{E}[Z] := (\mathrm{E}[Z_i])_{i=1,\dots,n} = \begin{bmatrix} \mathrm{E}[Z_1] \\ \vdots \\ \mathrm{E}[Z_n] \end{bmatrix} \tag{2.5}$$

and the covariance matrix *is defined as*

$$\begin{aligned} \mathrm{Cov}[Z, Z] := \left(\mathrm{Cov}[Z_i, Z_j]\right)_{i,j=1,\dots,n} &= \begin{bmatrix} \mathrm{Cov}[Z_1, Z_1] & \dots & \mathrm{Cov}[Z_1, Z_n] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}[Z_n, Z_1] & \dots & \mathrm{Cov}[Z_n, Z_n] \end{bmatrix} \\ &= \mathrm{E}\left[(Z - \mathrm{E}[Z])(Z - \mathrm{E}[Z])^{\mathrm{T}}\right]. \end{aligned} \tag{2.6}$$

With these definitions at hand, one can define the multivariate normal distribution, [22, Def. 3.1, p. 25].

Definition 2.3. *A vector* $X = (X_1, \dots, X_n)^{\mathrm{T}}$ *of real-valued random variables with expected value* $\mathrm{E}(X) = \mu$ *and positive definite covariance matrix* $C = \mathrm{Cov}[X, X]$ *is called* multivariate normally distributed*, provided that the corresponding density function is given by*

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} \exp\left(-\frac{1}{2}(x - \mu)^{\mathrm{T}} C^{-1} (x - \mu)\right). \tag{2.7}$$

Closely related to the framework presented in the next sections is the term of *Gaussian processes*, [71, § 2.3.2, p. 27], [51, p. 22, p. 27].

Definition 2.4. *Let* $\mathcal{D} \subset \mathbb{R}^d$ *and* $x \in \mathcal{D}$*. A family of random variables*

$$Y \equiv \{Y(x) \mid x \in \mathcal{D}\}$$

defined on a common probability space is called a Gaussian process*, if for any* $n \geq 1$ *and* *any choice of points* $x_1, \dots, x_n \in \mathcal{D}$ *the random vector*

$$Z := \begin{bmatrix} Y(x_1) \\ \vdots \\ Y(x_n) \end{bmatrix}$$

is multivariate normally distributed.

If the density function of a random variable or a random vector depends on an unknown parameter $\vartheta \in \Theta$, a popular approach to estimate this parameter is the method of maximum likelihood. Given a set of realizations of the random variable, the idea is to determine the parameter ϑ such that the likelihood to obtain the observed realizations is maximized, [45, § 13.1, p. 163].

Definition 2.5. *Let* $Y = (Y_1, \dots, Y_n)$ *be a vector of real-valued random variables of which the density function is given by* $f(\,\cdot \mid \vartheta)$ *with unknown parameter* $\vartheta \in \Theta$*. Furthermore let* $y = (y_1, \dots, y_n)$ *be a realization of this random vector. The function* L *with*

$$\vartheta \mapsto L(\vartheta) = f(y \mid \vartheta) \tag{2.8}$$

is called likelihood function *with respect to the realization* y*. The parameter which maximizes the likelihood function,*

$$\hat{\vartheta} := \operatorname*{arg\,max}_{\vartheta \in \Theta} L(\vartheta), \tag{2.9}$$

is called maximum likelihood estimate *of the parameter* ϑ*.*

2.2. Positive definite functions

The following Definition is based on [3, p. 67, Def. 1.1].

Definition 2.6. *Let* X *be a nonempty set and* $\phi\colon X \times X \to \mathbb{C}$*.*

(i) *The function* ϕ *is called a* positive definite kernel, *if and only if for an arbitrary set of* $n \in \mathbb{N}$ *mutually distinct points* $\{x^1, \dots, x^n\} \subseteq X$ *and arbitrary complex numbers* $\{\gamma_1, \dots, \gamma_n\} \subset \mathbb{C}$*, it holds*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_i \overline{\gamma_j} \phi(x^i, x^j) \geq 0. \tag{2.10}$$

(ii) The function ϕ is called a negative definite kernel, *if and only if it is hermitian, i.e. $\phi(x,y) = \overline{\phi(y,x)}$ for all $x, y \in X$, and for any set of $n \in \mathbb{N}$ mutually distinct points $\{x^1, \dots, x^n\} \subseteq X$ and complex numbers $\{\gamma_1, \dots, \gamma_n\} \subset \mathbb{C}$ with $\sum_{i=1}^{n} \gamma_i = 0$, it holds*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_i \overline{\gamma_j} \phi(x^i, x^j) \leq 0. \tag{2.11}$$

(iii) If the expression in Eq. (2.10) (resp. Eq. (2.11)) is zero only if $\gamma_i = 0, i = 1, \dots, n$, the kernel ϕ is called strictly positive definite *(resp.* strictly negative definite*).*

Note in particular that the expressions on the left hand sides of Eq. (2.10) and (2.11) need to be real-valued in order to satisfy the conditions. One can show that the above definition (i) of a positive definite kernel already implies that the kernel is hermitian, i.e. $\phi(x,y) = \overline{\phi(y,x)}$ for all $x, y \in X$, [3, p. 66].

As the name suggests, the term of positive definite kernels is closely related to the positive definiteness of quadratic matrices, [3, p. 67, Remark 1.2–1.3]:

Lemma 2.7. *The kernel $\phi\colon X \times X \to \mathbb{C}$ is positive definite, if and only if for any finite set of pairwise distinct points $\{x^1, \dots, x^n\} \subset X$, the matrix*

$$\Phi = \Big(\phi(x^i, x^j)\Big)_{i,j=1,\dots,n} \tag{2.12}$$

is positive semi-definite.

Note that the term *positive definite* kernel is a little misleading, since it is related to the *positive semi-definiteness* of quadratic matrices, cf. [3, p. 66].

In the last subsection, Gaussian processes were introduced as their modeling will play an important role in this work. The following theorem allows to set up a Gaussian process on an arbitrary real-valued positive definite function. It is taken from [51, Thm. 2, p. 27].

Theorem 2.8. *Let X be a nonempty set and $\phi\colon X \times X \to \mathbb{R}$ be a positive definite kernel. Then there exists a probability space and a Gaussian process defined on this probability space, whose covariance function is ϕ.*

Thus, real-valued kernels are of particular interest for this work. The next lemma states that it is sufficient for real-valued kernels to claim the analogue terms of Definition 2.6 on the real numbers, with the additional property of symmetry, [3, p. 68, Remark 1.6]:

Lemma 2.9. *Let X be a nonempty set. A real-valued kernel $\phi\colon X \times X \to \mathbb{R}$ is positive definite (resp. negative definite) in the sense of Definition 2.6, if and only if ϕ is symmetric, i.e. $\phi(x,y) = \phi(y,x)$ for all $x, y \in X$ and*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j \phi(x^i, x^j) \geq 0 \qquad (\textit{resp.} \leq 0) \tag{2.13}$$

for any set of $n \in \mathbb{N}$ mutually distinct points $\{x^1, \dots, x^n\} \subseteq X$ and arbitrary real numbers $\{\alpha_1, \dots, \alpha_n\} \subset \mathbb{R}$ (resp. with $\sum_{i=1}^{n} \alpha_i = 0$ in addition).

Some important examples of positive and negative definite real-valued kernels are given in Example 2.10, cf. [3, p. 69, Remark 1.9–1.10].

Example 2.10. *Let X be a nonempty set.*

(i) The constant kernel $\psi\colon X \times X \to \mathbb{R}$, $(x,y) \mapsto c$ is negative definite for $c \in \mathbb{R}$ and, at the same time, positive definite for $c \geq 0$, since for mutually distinct $\{x_1, \dots, x_n\} \subset X$ and arbitrary $\alpha_1, \dots, \alpha_n \in \mathbb{R}$,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j c = c \cdot \left(\sum_{i=1}^{n} a_i\right) \cdot \left(\sum_{j=1}^{n} a_j\right) = c \cdot \left(\sum_{i=1}^{n} a_i\right)^2 .$$

(ii) If $\phi\colon X \times X \to \mathbb{R}$ is positive definite and $t > 0$, then

$$\psi\colon X \times X \to \mathbb{R}, \quad (x,y) \mapsto -t \cdot \phi(x,y) \tag{2.14}$$

is negative definite.

(iii) The kernel $\psi\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $(x,y) \mapsto (x-y)^2$ is negative definite, since it is symmetric and for arbitrary mutually distinct points $x_1, \dots, x_n \in \mathbb{R}$ and any $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ with $\alpha_1 + \cdots + \alpha_n = 0$ it holds,

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j (x_i - x_j)^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j (x_i^2 - 2x_i x_j + x_j^2) \\
&= 2\left(\sum_{i=1}^{n} \alpha_i x_i^2\right) \cdot \left(\sum_{j=1}^{n} \alpha_j\right) - 2\sum_{i=1}^{n} \alpha_i x_i \cdot \sum_{j=1}^{n} \alpha_j x_j \\
&= -2\left(\sum_{i=1}^{n} \alpha_i x_i\right)^2 \leq 0.
\end{aligned}$$

The following results are useful for the construction of positive definite kernels.

As for positive semi-definite matrices, the set of positive definite kernels on $X \times X$ is closed under pointwise multiplication, [3, p. 69, Thm. 1.12]:

Theorem 2.11. *Let* $\phi_1, \phi_2 \colon X \times X \to \mathbb{C}$ *be two positive definite kernels. Then,*

$$\phi_1 \cdot \phi_2 \colon X \times X \to \mathbb{C} \tag{2.15}$$

is also positive definite.

The same holds true for the tensor product of two positive definite kernels, [3, p. 70, Corollary 1.13]:

Theorem 2.12. *Let* $\phi_1 \colon X \times X \to \mathbb{C}$ *and* $\phi_2 \colon Y \times Y \to \mathbb{C}$ *be two positive definite kernels. Then,*

$$\phi_1 \otimes \phi_2 \colon (X \times Y) \times (X \times Y) \to \mathbb{C}, ((x_1, y_1), (x_2, y_2)) \mapsto \phi_1(x_1, x_2) \cdot \phi_1(y_1, y_2) \tag{2.16}$$

is positive definite as well.

As a direct consequence, the following holds true.

Corollary 2.13. *Let* $d \in \mathbb{N}$ *and let* $\phi_i \colon \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ *for* $i = 1, \dots, d$ *be positive definite kernels. Then,*

$$\psi \colon \mathbb{C}^d \times \mathbb{C}^d \to \mathbb{C}, (x, y) \mapsto \prod_{i=1}^{d} \phi_i(x_i, y_i) \tag{2.17}$$

is positive definite.

This result allows to define positive definite kernels on subsets $X \subset \mathbb{R}^d$ or $X \subset \mathbb{C}^d$ of multidimensional vector spaces as a product of positive definite kernels on the one-dimensional spaces $\mathbb{R}^1$ or $\mathbb{C}^1$, respectively.

An important class of positive definite kernels is given by the following theorem from [3, p. 74, Thm. 2.2] which is mainly due to Schoenberg [73, p. 531, Thm. 4]:

Theorem 2.14. *Let* X *be a nonempty set and* $\psi \colon X \times X \to \mathbb{C}$*. Then,*

$$\phi \colon X \times X \to \mathbb{C}, \quad (x, y) \mapsto \exp(-t \cdot \psi(x, y)) \tag{2.18}$$

is a positive definite kernel for all $t > 0$*, if and only if* ψ *is negative definite.*

This yields directly to an important class of positive definite kernels:

Example 2.15. *Choosing the negative definite kernel*

$$\psi\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}, (x, y) \mapsto (x - y)^2$$

from Example 2.10 (iii) in Theorem 2.14 and applying Corollary 2.13 yields the positive definiteness of the so-called Gaussian correlation kernel,

$$\phi_2\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, (x, y) \mapsto \prod_{i=1}^{d} \exp\left(-\theta_i |x_i - y_i|^2\right), \tag{2.19}$$

for all $\theta_i \in \mathbb{R}_+$, $i = 1, \ldots, d$. It belongs to the family of exponential correlation kernels

$$\phi_p\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, (x, y) \mapsto \prod_{i=1}^{d} \exp\left(-\theta_i |x_i - y_i|^p\right), \tag{2.20}$$

with $\theta_i \in \mathbb{R}_+$, which are positive definite for all $p \in \mathbb{R}$ with $0 < p \leq 2$, see [73, Corollary 3, p. 532], [82, § 2.1, p. 515], [81].

A general characterization of the class of continuous positive semi-definite functions is given by Bochner's Theorem, [10, p. 155]:

Theorem 2.16. *Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be continuous and $\phi\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, (x, y) \mapsto f(x - y)$. The kernel ϕ is positive definite, if and only if f is a Fourier transform of a non-negative finite-valued Borel measure.*

The theorem gives a direct instruction on how to construct positive definite kernels. It is therefore often used to define covariance functions which match desired properties for the modeling of Gaussian processes. For more details on the definition of covariance kernels for Gaussian processes based on Bochner's Theorem, the reader is referred to [71, § 2.3.3, p. 31 et seqq.], [51, § 5–6, p. 30 et seqq.] and [44, § 4.3, p. 270] and the references therein, e.g. [57].

2.3. Schur complements

Next, the *Schur complement* is introduced. It serves as an useful tool to examine block matrices, which will often appear in the next sections of this thesis. The following definition and lemmas are based on [31, Chapter 16.1–16.2].

Definition 2.17. *Given any block matrix of the form*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \tag{2.21}$$

(i) If D is invertible, the matrix $A - BD^{-1}C$ is called Schur complement *of D in M.*

(ii) If A is invertible, the matrix $D - CA^{-1}B$ is called Schur complement *of A in M.*

The following lemma allows to construct the inverse of a block matrix based on its Schur complement, cf. [31, p. 433].

Lemma 2.18. *Let M be a block matrix as in (2.21) and suppose A and the Schur complement $S := D - CA^{-1}B$ of A in M are invertible. Then the inverse of M reads*

$$M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{bmatrix}. \tag{2.22}$$

To verify whether a symmetric block matrix is positive definite, the following lemma based on [31, Proposition 16.1 and 16.2, p. 434] is very useful.

Lemma 2.19. *Let M be a symmetric block matrix of the form*

$$M = \begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix}. \tag{2.23}$$

The following properties hold:

(i) If C is invertible, then M is positive definite, if and only if C and the Schur complement of C in M are positive definite.

(ii) If A is invertible, then M is positive definite, if and only if A and the Schur complement of A in M are positive definite.

2.4. Singular value decomposition

In this section, the singular value decomposition (SVD) is addressed, which is, due to its properties, a fundamental tool in model order reduction (MOR). We start with the mathematical definition of the SVD, which is based on the following theorem, [32, Theorem 2.4.1, p. 76].

Theorem 2.20. *Let* $Y \in \mathbb{R}^{m\times n}$ *be a real matrix of rank* r. *Then there exist orthogonal matrices*

$$U = [u^1, \dots, u^m] \in \mathbb{R}^{m\times m}, \qquad V = [v^1, \dots, v^n] \in \mathbb{R}^{n\times n}, \tag{2.24}$$

such that

$$U^{\mathrm{T}} Y V = \Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right] \in \mathbb{R}^{m\times n}, \tag{2.25}$$

where $\sigma_1 \geq \cdots \geq \sigma_r > 0$.

Definition 2.21. *The values* $\sigma_1, \dots, \sigma_r$ *in* (2.25) *are referred to as the* singular values *and the vectors* $u^1, \dots, u^m$ *and* $v^1, \dots, v^n$ *in* (2.24) *are called* left *and* right singular vectors *of the matrix* Y, *respectively.*

Remark 2.22. *Note that the first* $r = \operatorname{rank}(Y)$ *left singular vectors* $u^1, \dots, u^r$ *form an orthonormal basis of the subspace* $\mathcal{U} \subset \mathbb{R}^m$ *spanned by the columns* $y^1, \dots, y^n$ *of* Y. *The corresponding basis representation of column* $y^i, i = 1, \dots, n$ *is*

$$y^i = (\sigma_1 v_i^1) u^1 + \cdots + (\sigma_r v_i^r) u^r, \tag{2.26}$$

cf. [32, Corollary 2.4.6].

A nice property of the SVD which motivates its utilization in model order reduction, is given by the Eckart-Young Theorem, [32, Theorem 2.4.8].

Theorem 2.23. *If* $k < r = \operatorname{rank}(Y)$ *and*

$$Y_k := \sum_{i=1}^{k} \sigma_i u^i (v^i)^{\mathrm{T}}, \tag{2.27}$$

then Y_k *is the best rank-*k*-approximation of* Y *in the spectral norm in the sense that* Y_k *minimizes the error norm* $\|Y - B\|_2$ *among all matrices* $B \in \mathbb{R}^{m\times n}$ *of rank* k.

More precisely, it holds,

$$\min_{\operatorname{rank} B = k} \|Y - B\|_2 = \|Y - Y_k\|_2 = \sigma_{k+1}. \tag{2.28}$$

3. Basics of fluid dynamics

In this thesis, surrogate models are developed and tested for their applicability in automotive aerodynamic shape optimization. To provide an overview on the physical context of this application as well as existing solution methods, which will be used to obtain sampled data, a brief introduction to fluid mechanics and computational fluid dynamics (CFD) is given in this section.

The governing equations which describe the motion of a fluid are the *Navier-Stokes Equations*. They can be derived by applying the fundamental principles of conservation of mass, momentum and energy to the fluid, which is described in Sec. 3.1. In general, there is no analytic solution to the Navier-Stokes Equations for technically relevant flow problems, so that they need to be solved numerically. However, even on modern high-performance computers a direct numerical simulation of these flow problems is often not feasible. Some approaches to obtain approximate numerical solutions are discussed in Sec. 3.2. In the last subsection of this section, Sec. 3.3, the drag coefficient is introduced, which is used in industry to quantify the aerodynamic drag of a body immersed in a fluid.

3.1. The Navier-Stokes Equations

In this subsection, the main principles of fluid dynamics which lead to the Navier-Stokes Equations are discussed. It is mainly based on [8, Chapter 2].

As a starting point, suppose a three-dimensional spatial domain Ω_0 is filled with a fluid and we aim to describe the motion, the pressure and the energy of the fluid. A basic assumption in fluid dynamics is that the density of fluid molecules in the considered domain is high enough so that the fluid can be viewed as a continuum. This means that even in an infinitesimally small sub-volume there remain some fluid molecules which allow us to define for example the mean velocity or the mean kinetic energy. As a result, the density ρ, the velocity v and the energy E can be seen as field quantities that are well-defined at every location $x \in \Omega_0$ and for any time $t > 0$.

It is clear that from a microscopic point of view, this does not hold true. However, this *continuum assumption* is known to yield accurate descriptions of most macroscopic phenomena, [16, p. 2].

A description of the motion of the fluid can be derived by applying the three fundamental laws of mechanics to a spatially fixed sub-volume $\Omega \subset \Omega_0$—the *control volume*:

(i) The conservation of mass.

(ii) The conservation of momentum.

(iii) The conservation of energy.

In the following, let $\Omega \subset \Omega_0$ be a spatially fixed control volume in Ω_0 with boundary $\partial\Omega$. Let $n \in \mathbb{R}^3$ denote the unit normal vector at a point at the boundary $\partial\Omega$, pointing out of the control volume, as illustrated in Figure 1. The mass of the control volume is defined as the volume integral of the density field ρ over Ω,

$$m = \int_\Omega \rho \,\mathrm{d}x.$$

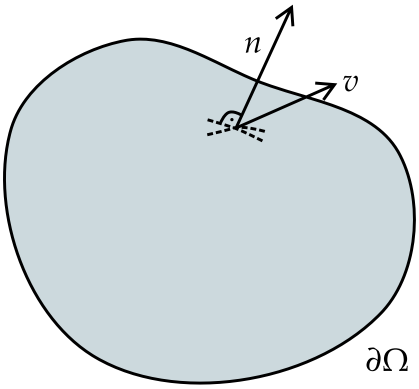

Figure 1.: Control volume Ω with boundary $\partial\Omega$ and surface normal n.

3.1.1. The conservation of mass

The principle of conservation of mass is based on the assumption that mass can neither be created nor destroyed. As a result, the change in mass in the control volume Ω equals the mass flux through its boundary $\partial\Omega$. Thus, it holds

$$\frac{\partial}{\partial t}\int_\Omega \rho \,\mathrm{d}x + \oint_{\partial\Omega} \rho v^{\mathrm{T}} n \,\mathrm{d}s = 0. \tag{3.1}$$

3.1.2. The conservation of momentum

Newton's second law states that the change in momentum of a body is caused by the net forces acting on the body. There are two groups of forces, which have to be considered: External body forces, which act on the mass of the control volume, for example gravity, and surface forces which act on the surface of the control volume, like pressure and friction, [8, p. 9]. If the body forces per unit volume are denoted by f_e, the pressure is denoted by p and the viscous stress tensor is denoted by $\tau \in \mathbb{R}^{3\times 3}$, the conservation of momentum can be formulated as

$$\frac{\partial}{\partial t}\int_{\Omega} \rho v \,\mathrm{d}x + \oint_{\partial\Omega} \rho v (v^{\mathrm{T}} n)\,\mathrm{d}s = \int_{\Omega} \rho f_e \,\mathrm{d}x - \oint_{\partial\Omega} p n \,\mathrm{d}s + \oint_{\partial\Omega} \tau n \,\mathrm{d}s. \tag{3.2}$$

The viscous stresses, which can be divided in normal and shear stresses, emerge from friction, [8, Chapter 2.3]. Air can be assumed to belong to the class of *Newtonian fluids*, [8, p. 13], for which the viscous stresses are proportional to the strain rate of the fluid. The shear stresses can thus be modeled as

$$\begin{aligned}
\tau_{12} = \tau_{21} &= \mu \left(\frac{\partial v_1}{\partial x_2} + \frac{\partial v_2}{\partial x_1} \right), \\
\tau_{13} = \tau_{31} &= \mu \left(\frac{\partial v_1}{\partial x_3} + \frac{\partial v_3}{\partial x_1} \right), \\
\tau_{23} = \tau_{32} &= \mu \left(\frac{\partial v_2}{\partial x_3} + \frac{\partial v_3}{\partial x_2} \right),
\end{aligned}$$

with the dynamic viscosity coefficient μ. When for the second viscosity coefficient λ the Stokes hypotheses $\lambda = -\frac{2}{3}\mu$ is used, [8, p. 15], the normal stresses can be modeled as

$$\begin{aligned}
\tau_{11} &= 2\mu \left(\frac{\partial v_1}{\partial x_1} - \frac{1}{3} \operatorname{div} v \right), \\
\tau_{22} &= 2\mu \left(\frac{\partial v_2}{\partial x_2} - \frac{1}{3} \operatorname{div} v \right), \\
\tau_{33} &= 2\mu \left(\frac{\partial v_3}{\partial x_3} - \frac{1}{3} \operatorname{div} v \right),
\end{aligned}$$

where $\operatorname{div} v = \frac{\partial}{\partial x_1} v_1 + \frac{\partial}{\partial x_2} v_2 + \frac{\partial}{\partial x_3} v_3$ denotes the divergence of the velocity field v.

3.1.3. The conservation of energy

The conservation of energy is justified by the first law of thermodynamics, which states that any changes in time of the total energy E of the control volume are due to the heat flux into the volume and the work done by the net forces, [8, Chapter 2.2.3]. The heat flux has two different sources: Diffusion due to temperature gradients ∇T

and heating from chemical reactions or radiation, for which we denote the time rate of heat flux per unit mass by $\dot{q}_h$. The net forces acting on the control volume can again be divided into the body forces f_e and the surface forces originating from pressure and friction, cf. Section 3.1.2. With the thermal conductivity coefficient k and absolute static temperature T, the equation for energy conservation reads

$$\begin{aligned}\frac{\partial}{\partial t}\int_{\Omega}\rho E\,\mathrm{d}x+\oint_{\partial\Omega}\rho E(v^{\mathrm{T}}n)\,\mathrm{d}s=\int_{\Omega}(\rho f_e^{\mathrm{T}}v+\dot{q}_h)\,\mathrm{d}x-\oint_{\partial\Omega}p(v^{\mathrm{T}}n)\,\mathrm{d}s&\\+\oint_{\partial\Omega}(\tau v)^{\mathrm{T}}n\,\mathrm{d}s+\oint_{\partial\Omega}k(\nabla T)^{\mathrm{T}}n\,\mathrm{d}s.&\end{aligned}\tag{3.3}$$

Using the total enthalpy,

$$H=E+\frac{p}{\rho},\tag{3.4}$$

Eq. (3.3) can be rewritten as

$$\begin{aligned}\frac{\partial}{\partial t}\int_{\Omega}\rho E\,\mathrm{d}x+&\oint_{\partial\Omega}\rho H(v^{\mathrm{T}}n)\,\mathrm{d}s\\&=\int_{\Omega}(\rho f_e^{\mathrm{T}}v+\dot{q}_h)\,\mathrm{d}x+\oint_{\partial\Omega}(\tau v)^{\mathrm{T}}n\,\mathrm{d}s+\oint_{\partial\Omega}k(\nabla T)^{\mathrm{T}}n\,\mathrm{d}s.\end{aligned}\tag{3.5}$$

3.1.4. The governing equations

The exchange of mass, momentum and energy through the boundary of a spatially fixed control volume Ω is completely described by the conservation equations (3.1), (3.2) and (3.5) presented in the previous subsections. They can be summarized by the following system of coupled partial differential equations, the so-called *Navier-Stokes Equations*,

$$\frac{\partial}{\partial t}\int_{\Omega}W\,\mathrm{d}x+\oint_{\partial\Omega}(F_c-F_v)\,\mathrm{d}s=\int_{\Omega}Q\,\mathrm{d}x,\tag{3.6}$$

where W is the vector of so-called *conservative variables*, F_c denotes the vector of convective fluxes, F_v is the vector of viscous fluxes and Q contains the source terms,

$$W=\begin{bmatrix}\rho\\\rho v\\\rho E\end{bmatrix},\quad F_c=\begin{bmatrix}\rho v^{\mathrm{T}}n\\\rho v(v^{\mathrm{T}}n)+pn\\\rho H(v^{\mathrm{T}}n)\end{bmatrix},\quad F_v=\begin{bmatrix}0\\\tau n\\\left((\tau v)^{\mathrm{T}}+k\nabla T\right)n\end{bmatrix},\quad Q=\begin{bmatrix}0\\\rho f_e\\\rho f_e v+\dot{q}_h\end{bmatrix},$$

cf. [8, Chapter 2.4]. The Navier-Stokes Equations consist of 5 coupled partial differential equations for 7 unknown flow field variables, which are the density ρ, the velocity $v=(v_1,v_2,v_3)^{\mathrm{T}}$, the energy E, the pressure p and the temperature T. Thus, additional information on the total energy and the thermodynamic relations between the state variables for the considered fluid is needed to close the system, [24, Chapter 10.2, p. 368]. In most aerodynamic applications it is reasonable to assume that air behaves like a *perfect gas*. This assumption yields the additional relations

$$p = \rho R T, \qquad p = (\gamma - 1)\rho \left(E - \frac{|v|^2}{2} \right), \tag{3.7}$$

where the specific gas constant $R = c_p - c_V$ is defined as the difference and $\gamma = \frac{c_p}{c_V}$ as the ratio of specific heat capacity at constant pressure c_p to specific heat capacity at constant volume c_V, [8, Chapter 2.4.1][4].

Remark 3.1. *If viscous effects are neglected, that is* $F_v = 0$ *in Eq.* (3.6)*, the Navier-Stokes Equations reduce to the Euler Equations, [8, p. 24].*

Remark 3.2. *If the Mach number, which is the ratio of the local velocity of the fluid to the ambient speed of sound, is less than* 0.3*, the flow can be assumed to be incompressible, [24, Chapter 1.7.1]. In that case, the conservation of mass yields* $\operatorname{div} v = 0$ *and the viscous stresses simplify, cf. Sec. 3.1.2. In car aerodynamics, the Mach number is typically below* 0.3 *and thus the assumption that the fluid is incompressible is often valid. For a discussion on the influence of compressibility in the context of car aerodynamics see [74, Chapter 2.4.2].*

3.2. Computational fluid dynamics

Suppose we aim to compute the flow field for a car in a windtunnel because we would like to study the aerodynamic behaviour of its shape. To model the flow around the vehicle the Navier-Stokes Equations (3.6), which mathematically describe the conservation of mass, momentum and energy for the fluid in the windtunnel, are complemented by some initial and boundary conditions. For example the velocity of the incoming flow needs to be specified and we need to demand that the fluid velocity directly at the surface of the vehicle and the walls of the wind tunnel is 0. Throughout this work, such Navier-Stokes initial and boundary value problems will be referred to as *flow problems*.

Except for some very simple test cases, there is no analytic solution to a Navier-Stokes initial and boundary value problem, [8, p. 29]—in the general case, it is not

---

[4] For air, the specific gas constant R was found to be $R = 287.2\,\frac{\mathrm{N\,m}}{\mathrm{kg\,K}}$ and the specific heat ratio at $0\,^\circ\mathrm{C}$ is $\gamma = 1.400$, [47, p. 323, Tab. T-3].

even clear if a solution exists at all[5]. Therefore, in general, flow problems need to be solved numerically on a discrete grid. However, technically relevant flow problems often feature turbulent flow profiles, meaning that fluid particles with different velocity and momentum mix as they move, [29, p. 41]. As a result, the corresponding flow field is highly complex and features small scale fluctuations. To perform an accurate direct numerical simulation (DNS) of the flow, these small scale fluctuations need to be resolved which calls for a very fine discretization in space and time. This in turn requires such high computational and memory capacities that even on modern high-performance computers DNS is not (yet) feasible for most industrial problems, [8, p. 53]. There exist however several approaches to overcome this problem. The approaches used in the numerical applications considered in this work are briefly described below.

3.2.1. Reynolds-averaged Navier-Stokes simulation

This subsection is based on a brief summary of Reynolds-averaged Navier-Stokes simulation given in [8, Chapter 3.3]. For a more detailed discussion, the reader is referred to [74, Chapter 14.1.4.1] or [24, 9.4].

Often one is more interested in time averaged quantities instead of the complete time history of the flow variables. For example, in the car industry, one might be interested in the averaged drag force of a car under certain conditions instead of the fluctuations of this quantity due to turbulent flows. The idea of Reynolds-averaged Navier-Stokes (RANS) simulations is to decompose the flow variables in a time independent mean and a turbulent, fluctuating-in-time part followed by time or ensemble averaging. Applying the averaging to the Navier-Stokes Equations yields governing equations for the mean variables which have basically the same structure as the original Navier-Stokes Equations but feature two additional terms: the so-called *Reynolds-stress tensor* and the *turbulent heat flux vector*. The new variables can not be resolved by the new governing equations alone and therefore a modeling of these quantities is required. This is done by so-called *turbulence models*. There exists a variety of different turbulence models. An overview of different approaches is given in [24, Chapter 9.4.2], [74, Chapter 14.1.4.1.2] and [8, Chapter 3.3].

Compared to DNS and the large-eddy simulation, which is presented in the next subsection, the RANS simulation has the advantage that a considerably coarser grid

---

[5] The problem of existence and uniqueness of solutions to the Navier-Stokes Equations belongs to the 7 famous “Millennium Problems” of the *Clay Mathematics Institute*, [17].

can be taken and the mean quantities can be assumed to be stationary, which significantly reduces the computational effort, [8, p. 54]. On the other hand, RANS simulations do not resolve the details of the turbulent flow structure.

3.2.2. Large-eddy simulation

A short summary of the idea of Large-eddy simulations is given in this subsection. It is based on [74, Chapter 14.1.4.2].

The principle of large-eddy simulations (LES) is based on the observation that in turbulent flows large and small scale flow vortices take different roles in the flow field. While large scale eddies are responsible for the bulk of energy and momentum transport, small scale vortices are more important to friction-related dissipation of motion into heat energy. Furthermore, small scale vortices feature a more universal character than large scale eddies, meaning that their structure is similar for many flow problems. The idea is therefore to approximate the small scale vortices by a suitable model instead of resolving them in terms of the Navier-Stokes Equations. To this end, the flow variables are decomposed into a large and a small scale part by applying a mathematical filter (e.g. a Gauss filter). Similar to RANS, the filtering introduces additional unknowns into the governing equations, the so-called *subgrid-scale stresses* which are then modeled via *subgrid-scale models*.

To resolve the high-frequency fluctuations in the flow field, LES requires considerably smaller time step sizes than RANS. In addition, especially in wall regions, a very fine grid is needed to accurately capture the motion of the fluid. As a result, LES is much more computationally demanding than RANS simulations.

3.2.3. Detached-eddy simulation

This Subsection is based on [74, Chapter 14.1.4.2.3].

Due to the requirement of extremely fine grids in near-wall regions, LES might become unfeasible for industrial applications. This issue associated with LES is addressed by detached-eddy simulations (DES), which combine LES with RANS. Close to walls RANS simulation with a suitable turbulence model is used. In all other regions, LES is run to simulate the motion of the fluid. This allows to use a coarser grid in near-wall regions compared to LES and thus the computational effort and the memory requirements are reduced. However, the second issue associated with

LES, the requirement of small time step sizes, remains, so that DES is still computationally very demanding.

3.3. The drag coefficient

In car aerodynamics, the drag force F_d plays an important role for emission and consumption and is thus of particular interest, [74, §3.2.2, p. 141 et seq.]. It is the component of the fluid dynamic force which acts opposite to the direction of travel. More precisely, if $\partial\Omega'$ defines the boundary of the vehicle and $x_d \in \mathbb{R}^3$ is the unit vector in the direction of travel, the drag force is given by $F_d = x_d^{\mathrm{T}} F$, where

$$F = \oint_{\partial\Omega'} \tau n \, \mathrm{d}s - \oint_{\partial\Omega'} pn \, \mathrm{d}s, \tag{3.8}$$

is the vector of fluid dynamic forces acting on the surface of the vehicle, cf. the surface forces in Eq. (3.2). The drag force is strongly dependent on the projected frontal area A of the vehicle and its velocity v_∞. A dimensionless quantity which is usually used to quantify the drag of a body shape is the *drag coefficient*. It is defined as

$$c_d = \frac{2F_d}{\rho v_\infty^2 A}. \tag{3.9}$$

4. Constructing a surrogate model

The construction of a surrogate model is often motivated by the fact that an evaluation of the original model is computationally expensive and it is thus not feasible—in some cases even impossible—to use it in optimization processes. In contrast to the evaluation of the original high-fidelity model, where the computational time is (more or less) proportional to the number of evaluations, the surrogate model needs to be constructed before it can be used as an approximation to the original model. Therefore, in the context of surrogate modeling, one always distinguishes between the so called *offline* and the *online phase* of a surrogate modeling process. The offline phase contains all steps which need to be done until the surrogate model can be evaluated at arbitrary parameter configurations. In comparison, the online phase consists of every computation which needs to be performed to evaluate a ready-made surrogate model. Figure 2 qualitatively describes the relationship between the computational time and the number of evaluations for the original and the surrogate model.

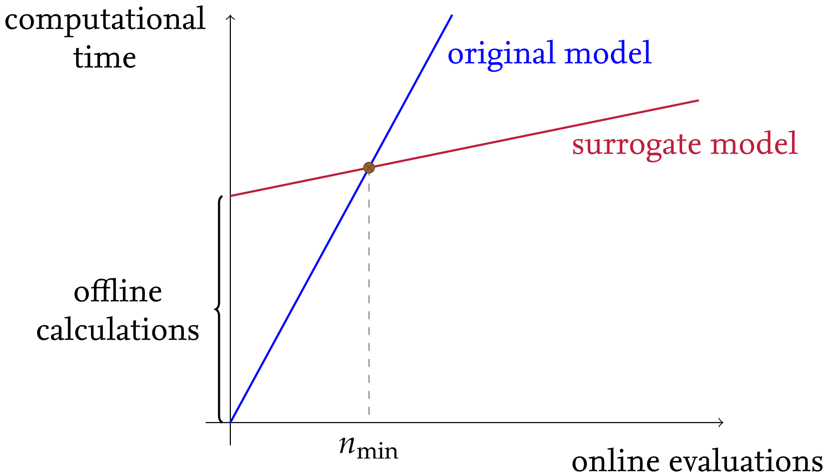

Figure 2.: Comparison of the computational effort of a surrogate model with online and offline computations and the original high-fidelity model.

From the figure it is clear that the construction of a surrogate model does only pay off, when the number of online evaluations is large enough. The critical number of online evaluations n_{min} is not only dependent on the computational time of an online evaluation, but is also strongly influenced by the duration of the offline phase. Beside a fast evaluation time of a surrogate model, it is therefore also important to keep the time for offline computations short. The typical offline phase of data-fit surrogate modeling techniques, which are in the focus of this work, is displayed in a flow chart in Figure 3. The approaches are fully data-driven, i.e. the original model is viewed as a black box, mapping some input variables to a scalar- or vector-valued output. Sample observations of input-output-pairs of the original model are used to compute an approximation to the original model based on interpolation. The surrogate modeling process—from the original model to the surrogate model—is described in the following.

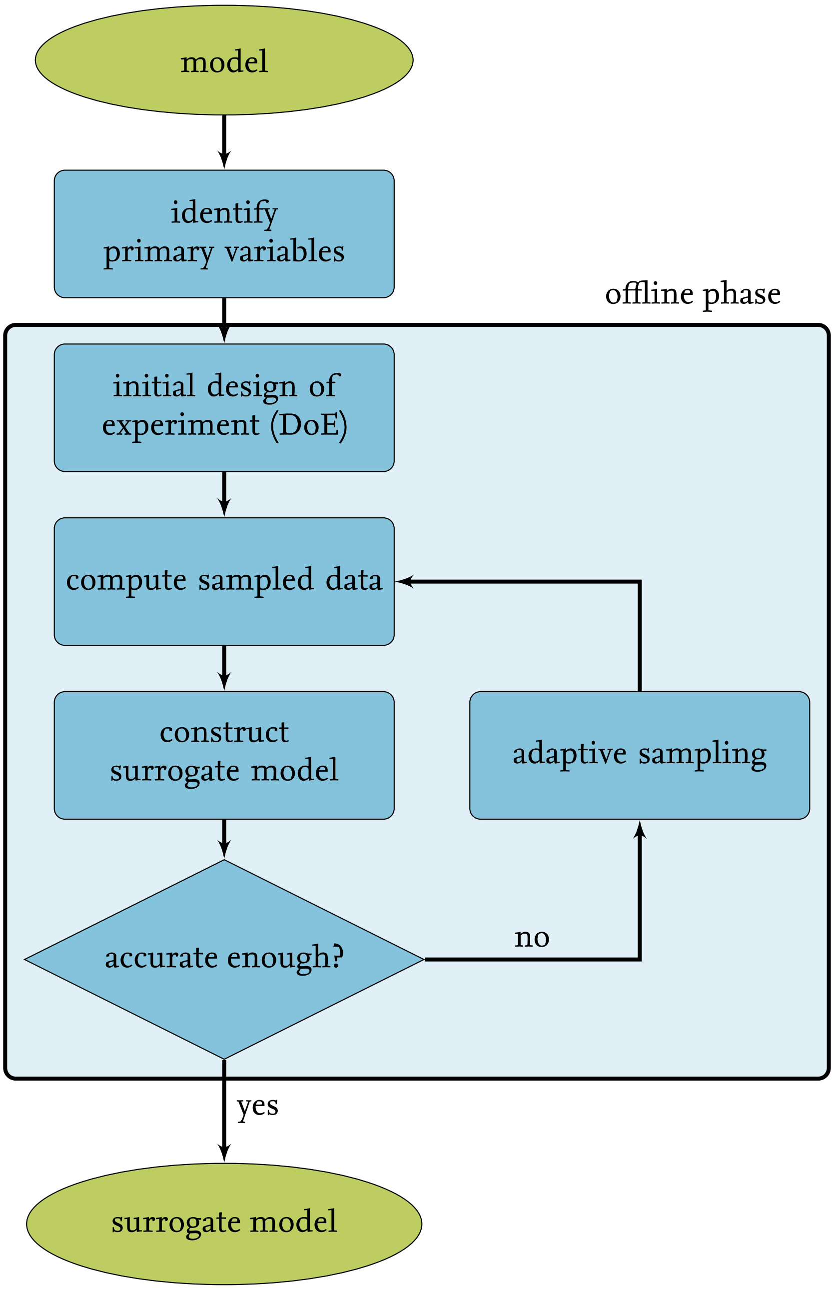

Figure 3.: Steps of constructing a surrogate model

The first step of constructing a surrogate model is the identification of primary variables. The original model may be dependent on a large number of input variables. For example, consider the problem of car shape optimization. At least in theory, the shape of a vehicle can be changed continuously and arbitrarily and the original model is able to handle arbitrary vehicle shapes. But this would lead to an infinite number of input parameters, which is not practicable—neither in numerical optimization processes, nor with the purpose of surrogate modeling. The shape needs to be parametrized and optimization parameters need to be introduced purposefully. The more parameters are introduced, the more complex the optimization problem becomes and the more sample points are needed to build an accurate surrogate model. Suppose we have sampled a one-dimensional input space $n = 10$ times to obtain a certain level of accuracy of the surrogate model. If we increase the number of input parameters to $k = 3$, we need to compute $n^k = 1000$ sample points to gain the same sample density[6]. Since the evaluation of the original model is computationally expensive, this clearly has a huge impact on the offline phase of the surrogate modeling process. Thus, the number of input parameters has to be reduced to a manageable amount.

After the desired input parameters are determined, a *design of experiment* has to be chosen. To obtain a good surrogate model, the sampling points should be chosen in a way that the characteristics of the response are captured by the samples. If hardly anything is known about the response, it is impossible to know in advance, whether a chosen sampling plan captures the characteristics or not. By using *adaptive sampling* strategies the accuracy of a surrogate model can be controlled and sample points are chosen iteratively at locations, where the estimated error of the surrogate model is large. As an initial sample set, a full factorial design or a random latin hypercube design with a small number of initial sample points can be used, e.g. [28, p. 13 et seq.].

In applications, where the computational time for the evaluation of the original model is extremely high, the maximum number of sample points is determined by the computational time one can afford.

When the initial sample set is chosen, the original model is evaluated at these sites to obtain the corresponding responses. For example, in the industrial application presented in this work, a quantity of interest will be a distribution of a desired flow quantity in a steady time-independent flow field around the vehicle's body, e.g. the

---

[6] This problem is known as the *curse of dimensionality*, see e.g. [28, Chapter 1.1]

pressure distribution on the car's shape or the flow velocities in the wind tunnel domain. The input-output pairs are then interpolated to get an initial surrogate model. Details on this modeling process are given in Section 5 and Section 6.

The quality of the surrogate model can be assessed by using the well-known *leave-one-out cross-validation*, see Sec. 6.4.1, and the error estimator provided by the considered interpolation method, see Remark 5.1, Remark 5.2 and Sec. 6.4. As long as a pre-defined stopping criterion is not satisfied and the maximum number of sample points is not reached, new sample points are added adaptively at sites, where the estimated error of the surrogate model is highest. The corresponding responses are computed at these sites, a new surrogate model is constructed and its accuracy is assessed.

After this loop is left, the surrogate model enters the online phase and is now ready for evaluation.

5. Surrogate modeling for scalar-valued responses

In this section, we focus on data-driven surrogate modeling techniques for scalar-valued responses. We start by giving an overview of the most popular methods. Suppose there is a functional relationship between the scalar-valued quantity of interest y and a number of input variables $x \in \mathcal{D} \subset \mathbb{R}^d$. The common goal of the presented methods is to find a mapping $\hat{y} \colon \mathcal{D} \to \mathbb{R}$ of a simple mathematical structure between the input variables and the quantity of interest, which matches observations of input-output-pairs $(x^i, y(x^i)), i = 1, \dots, n$ of the original model, i.e. $\hat{y}(x^i) = y(x^i), i = 1, \dots, n$.

In the *polynomial response surface method*, [59], [28, § 2.2, p. 40 et seqq.], the observed data is interpolated by a low-order polynomial, that is a linear combination of multivariate monomials. A difficulty associated with these models is that the selection of the order of the polynomial as well as the monomials and corresponding coefficients is sophisticated. Furthermore, due to their relatively simple structure, they are not suitable to approximate complex mappings as they appear in many real-life applications, [28, p. 75].

A more flexible approach is given by *radial basis function* models (RBF), [10], [28, § 2.3, p. 45 et seqq.], where the original model is approximated by a linear combination of radially symmetrical basis functions. While the computation of the corresponding basis coefficients reduces to the solution of a linear system of equations, radial basis function models can handle even highly nonlinear responses, [28, p. 46].

The same holds true for the *Kriging* method, [28, § 2.4, p. 49 et seqq.], where the original model is viewed as a realization of a random process. A surrogate model is then constructed to be the best linear unbiased estimator to this random function on the basis of the observed data. Besides its flexibility, this approach has the advantage that it comes with a mean squared error estimator, which can be used to adaptively improve the surrogate model by including additional sampled data. However, as with radial basis function models and any other interpolation technique, the computation of accurate Kriging models requires sufficiently many sample points. Especially if the number of input parameters is large and the evaluation of the original model is computationally expensive, this task becomes at least demanding.

In this situation, *variable-fidelity surrogate models* are a promising alternative if in addition to the computationally expensive original model (at least) a second model is available, which describes the same quantity, is cheaper to evaluate but less accurate. Samples of this lower-fidelity model are then used to obtain the global trend of the quantity of interest in the entire domain, and in this way improve the interpolation of the high-fidelity data. As a result, fewer high-fidelity sample points are needed than for the single-fidelity approaches. Thus, the efficiency of constructing a surrogate model is increased.

The simplest way to incorporate data of lower fidelity is by using a so-called *bridge function*, which aims to correct a surrogate model for the low-fidelity data to interpolate the high-fidelity data. Depending on the type of the correction, the bridge functions are either of additive, multiplicative or hybrid type, [35, § 4, p. 181 et seq.].

A direct extension of the Kriging method to handle variable-fidelity data goes under the name *Cokriging*, [28,36,40]. The high-fidelity model and the low-fidelity model are viewed as realizations of correlated random processes and the best linear unbiased estimator is constructed based on the observed data. Another Kriging-based variable-fidelity modeling approach which is easier to implement than the Cokriging method is *Hierarchical Kriging*, [34]. Both methods have been successfully applied in aerodynamic applications, [27,34,36]. As with Kriging, the methods Cokriging and Hierarchical Kriging provide a mean squared error estimator.

Since their ability to handle even highly nonlinear responses and the advantage of providing a useful error estimate, we focus on Kriging-based surrogate models in this work. The Kriging approach is presented in detail in Sec. 5.1. In Sec. 5.2, the variable-fidelity method Cokriging is described, followed by a discussion on open theoretical questions associated with the technique. Subsequently, the variable-fidelity method Hierarchical Kriging is presented in Sec. 5.3.

5.1. Kriging

Kriging, sometimes also called Gaussian process regression (GPR), the spatial linear model or spatial Gaussian process, is a statistical interpolation method, based on the Gauss-Markov Theorem, [45, p. 172]. A method similar to what is now known as Kriging was applied in a geostatistic context by Danie Krige in 1951 [46]. In 1963, Georges Matheron [55] specified the method mathematically and introduced the term “Krigage” to honor Krige, [19]. Since then, many variations and extensions of the original model have been published by different authors. Some examples are Gradient-Enhanced Kriging, [58] and the variable-fidelity surrogate modeling techniques Cokriging, e.g. [28, 36, 40], and Hierarchical Kriging, [34].

Among different other applications, Kriging-based methods gain popularity in the context of design and analysis of computer experiments, [28, 44, 53, 70, 71]. Kriging methods allow to construct efficient emulators for accurate but computationally expensive computer models with scalar-valued outputs—even if the response surface has a highly nonlinear structure.

For the sake of completeness of this thesis, this section gives an introduction to the mathematical formulation of Kriging, based on the work of Sacks et al. [70], Santner et al. [71] and Forrester et al. [28, § 2.4].

5.1.1. Basic model assumptions of Kriging

A basic assumption of the Kriging framework is that the true functional relationship $y\colon \mathcal{D} \to \mathbb{R}$ between the input variables $x \in \mathcal{D} \subset \mathbb{R}^d$ and their corresponding scalar-valued output $y(x) \in \mathbb{R}$, is a realization of a random function,

$$\mathcal{Y}(x) = f(x)\beta + \varepsilon(x), \tag{5.1}$$

where $f\colon \mathbb{R}^d \to \mathbb{R}^p$ is a known regression model with corresponding unknown regression parameters $\beta \in \mathbb{R}^p$ and $\varepsilon(x)$ is a Gaussian process with expected value $\mathrm{E}[\varepsilon(x)] = 0$ and stationary variance $\mathrm{Var}[\varepsilon(x)] = \sigma^2 > 0$. Further it is presumed that the covariance of the Gaussian process $\varepsilon(x)$ is given by a known stationary spatial correlation kernel $\mathcal{R}_\theta(x, \hat{x})$,

$$\mathrm{Cov}[\epsilon(x), \epsilon(\hat{x})] = \sigma^2 \mathcal{R}_\theta(x, \hat{x}), \tag{5.2}$$

for $x, \hat{x} \in \mathcal{D}$ with (usually) unknown correlation parameters $\theta \in \mathbb{R}^d_+$. While this approach to model the covariances via a spatial correlation kernel is common in the context of design and analysis of computer experiments, [28, 53, 70], there exist other accesses like the so-called *variogram approach* used by the geostatistic community, see e.g. [41, p. 31 et seq.].

Because it is assumed that the error $\varepsilon(x)$ is a Gaussian process, the chosen correlation kernel is required to yield a positive definite covariance matrix

$$C := \left(\sigma^2 \mathcal{R}_\theta(x^i, x^j)\right)_{i,j=1,\dots,n}, \tag{5.3}$$

for any choice of pairwise distinct sample points $x^1, \dots, x^n \in \mathcal{D}$, cf. Section 2.1. Therefore positive definite functions are particularly suitable, since these functions guarantee that the covariance matrix is positive definite for mutually distinct input points, cf. Lemma 2.7. A common approach for the construction of valid correlation kernels is based on Bochner's Theorem 2.16. A review of this theory is beyond the scope of this work. The reader is referred to [71, § 2.3.3, p. 31 et seqq.], [20, p. 84], [21, p. 956 et seqq.] and [57] for more details.

Some popular choices of spatial correlation kernels, which are all constructed based on Bochner's Theorem, see e.g. [71, Example 2.3, Example 2.4], [44, § 4.3.2., p. 472] [57, Corollary 2, p. 112], are listed in Table 1.

Correlation Kernel	$\mathcal{R}_\theta(x, \hat{x})$
exponential	$\prod_{i=1}^{d} \exp(-\theta_i \lvert x_i - \hat{x}_i \rvert)$, where $\theta_i \in \mathbb{R}_+$
Gaussian	$\prod_{i=1}^{d} \exp(-\theta_i \lvert x_i - \hat{x}_i \rvert^2)$, where $\theta_i \in \mathbb{R}_+$
cubic	$\prod_{i=1}^{d} (1 - 3\xi_i^2 + 2\xi_i^3)$, where $\xi_i = \min\{1,\ \theta_i \lvert x_i - \hat{x}_i \rvert\}$ and $\theta_i \in \mathbb{R}_+$

Table 1.: Some popular choices of the correlation kernel, [53]

For comparison, the three different correlation kernels of Table 1 are displayed in Figure 4 for the one dimensional case $d = 1$, and fixed correlation parameter $\theta = 1.0$.

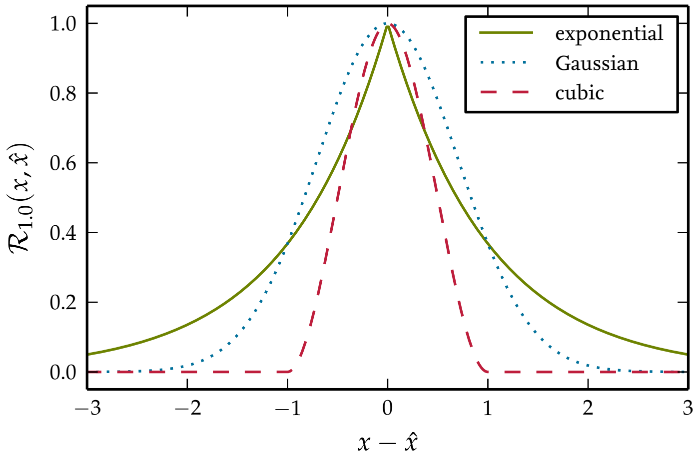

Figure 4.: The exponential, Gaussian and cubic correlation kernel for the one dimensional case and fixed correlation parameter $\theta = 1.0$

The chosen correlation kernel should match the properties of the functional response/the physical phenomenon, which should be approximated by the Kriging model. For example, if the response is expected to be smooth and differentiable, the correlation function will most likely feature a parabolic behavior near the origin, as the Gaussian or the cubic correlation kernel. On the other hand, if the response is expected to be irregular, a non-differentiable correlation kernel, such as the exponential kernel, might perform better, [53, p. 7], [70, p. 413].

The correlation parameters $\theta = [\theta_1, \ldots, \theta_d]$ act like distance weights, as Figure 5 illustrates. The correlation of two fixed distinct points $x, \hat{x} \in \mathbb{R}^d$ increases as the parameters θ_i decrease. The choice of the correlation parameter θ has a large impact on the modeled correlation and thus on the Kriging model itself. Therefore, this parameter is usually fitted by a maximum likelihood optimization, [70, p. 413], [28, pp. 53]. Details will be given in Section 5.1.3.

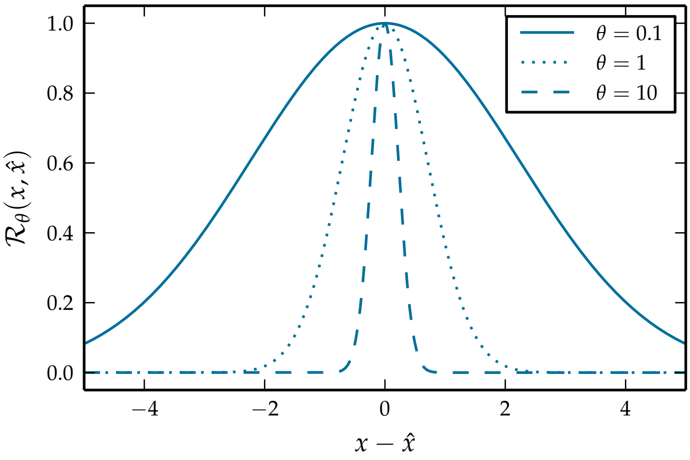

Figure 5.: One dimensional Gaussian correlation kernel for different choices of the correlation parameter θ

Typical choices of the regression model are

- constant regression ("Ordinary Kriging")
  $f\colon \mathbb{R}^d \to \mathbb{R},\ x \mapsto 1$ with $\beta \in \mathbb{R}$
- linear regression ("Universal Kriging")
  $f\colon \mathbb{R}^d \to \mathbb{R}^{d+1},\ x \mapsto (1, x_1, \ldots, x_d)$ with $\beta = (\beta_0, \ldots, \beta_d)^{\mathrm{T}} \in \mathbb{R}^{d+1}$.

5.1.2. Construction of the Kriging predictor

Suppose the function $y(x)$ is sampled at n sample locations

$$\mathcal{X} = \{x^1, \dots, x^n\} \subset \mathbb{R}^d.$$

The corresponding outputs are given by

$$Y = [y^1, \dots, y^n]^{\mathrm{T}} := [y(x^1), \dots, y(x^n)]^{\mathrm{T}} \in \mathbb{R}^n. \tag{5.4}$$

The Kriging predictor is constructed such that it is the best linear unbiased estimator (BLUE) to the observed data. Let

$$F := \begin{bmatrix} f(x^1) \\ \vdots \\ f(x^n) \end{bmatrix} \in \mathbb{R}^{n \times p}, \tag{5.5}$$

denote the matrix of regression vectors and $\Sigma := Y - F\beta \in \mathbb{R}^n$ the corresponding vector of errors to the observed function values, respectively. The Kriging predictor $\hat{y}(x)$ is constructed to be linear in the observed data (“L” in BLUE),

$$\hat{y}(x) = \sum_{i=1}^{n} \lambda_i(x) y^i = \lambda^{\mathrm{T}}(x) \cdot Y = \lambda^{\mathrm{T}}(x)(F\beta + \Sigma), \tag{5.6}$$

with weights $\lambda(x) = (\lambda_1(x), \dots, \lambda_n(x))^{\mathrm{T}} \in \mathbb{R}^n$. The weights are computed such that the Kriging estimator is the best predictor (“B” in BLUE) in the sense that it minimizes the mean squared error among all unbiased estimators (“U” in BLUE).

The unbiasedness constraint reads

$$\begin{aligned} 0 = \mathrm{E}\left[\hat{y}(x) - y(x)\right] &= \mathrm{E}\left[\lambda^{\mathrm{T}}(x)\,(F\beta + \Sigma) - (f(x)\beta + \varepsilon(x))\right] \\ &= \mathrm{E}\left[\lambda^{\mathrm{T}}(x)\Sigma + \left(\lambda^{\mathrm{T}}(x)F - f(x)\right)\beta - \varepsilon(x)\right] \\ &= \lambda^{\mathrm{T}}(x)\,\underbrace{\mathrm{E}[\Sigma]}_{=0} + \left(\lambda^{\mathrm{T}}(x)F - f(x)\right)\beta - \underbrace{\mathrm{E}[\varepsilon(x)]}_{=0} = \left(\lambda^{\mathrm{T}}(x)F - f(x)\right)\beta \end{aligned}$$

and thus for arbitrary regression models

$$\lambda^{\mathrm{T}}(x)F = f(x). \tag{5.7}$$

For every untried location $x \in \mathcal{D}$, the estimator $\hat{y}$ is computed such that the mean squared error (MSE) is minimized,

$$\begin{aligned} \mathrm{MSE}(x) &= \mathrm{E}\left[(\hat{y}(x) - y(x))^2\right] = \mathrm{E}\left[\left(\lambda^{\mathrm{T}}\Sigma + \left(\lambda^{\mathrm{T}}F - f(x)\right)\beta - \varepsilon(x)\right)^2\right] \\ &= \mathrm{E}\left[\lambda^{\mathrm{T}}\Sigma\Sigma^{\mathrm{T}}\lambda - 2\lambda^{\mathrm{T}}\Sigma\varepsilon(x) + \varepsilon(x)^2\right] \\ &= \lambda^{\mathrm{T}}(x)\,\mathrm{E}\left[\Sigma\Sigma^{\mathrm{T}}\right]\lambda - 2\lambda^{\mathrm{T}}\,\mathrm{E}\left[\Sigma\varepsilon(x)\right] + \mathrm{E}\left[\varepsilon(x)^2\right] \\ &= \sigma^2\lambda^{\mathrm{T}}R\lambda - 2\sigma^2\lambda^{\mathrm{T}}r(x) + \sigma^2, \end{aligned} \tag{5.8}$$

where $\lambda \equiv \lambda(x)$,

$$R := \mathrm{E}\left[\Sigma\Sigma^{\mathrm{T}}\right] = \left(\mathrm{Cor}(\varepsilon(x^i), \varepsilon(x^j))\right)_{i,j=1,\ldots,n} = \left(\mathcal{R}_\theta(x^i, x^j)\right)_{i,j=1,\ldots,n}$$

is the correlation matrix and

$$r(x) := \mathrm{E}\left[\Sigma\varepsilon(x)\right] = \left(\mathrm{Cor}(\varepsilon(x^i), \varepsilon(x))\right)_{i=1,\ldots,n} = \left(\mathcal{R}_\theta(x^i, x)\right)_{i=1,\ldots,n}$$

is the vector of correlations of the sample locations and the location x.

Thus, for every location $x \in \mathcal{D}$, an optimal choice of weights $\lambda(x)$ is a solution of the following constrained optimization problem,

$$\min_{\lambda(x)} \sigma^2\left(\lambda^{\mathrm{T}}(x)R\lambda(x) - 2\lambda^{\mathrm{T}}(x)r(x) + 1\right), \tag{5.9}$$

$$\text{s.t. } \lambda^{\mathrm{T}}(x)F = f(x). \tag{5.10}$$

By introducing Lagrange multipliers $\mu \in \mathbb{R}^p$, this constrained optimization problem can be reformulated into an unconstrained one,

$$\min_{\lambda=\lambda(x)} \underbrace{\sigma^2\left(\lambda^{\mathrm{T}}R\lambda - 2\lambda^{\mathrm{T}}r(x) + 1\right) + \left(\lambda^{\mathrm{T}}F - f(x)\right)\mu}_{\psi(\lambda)}. \tag{5.11}$$

The first partial derivative of the objective function $\psi(\lambda)$ with respect to λ is

$$\frac{\partial}{\partial\lambda}\psi(\lambda) = 2\sigma^2\left(R\lambda - r(x) + F\frac{\mu}{2\sigma^2}\right) \tag{5.12}$$

and the corresponding Hessian is

$$H = 2\sigma^2 R. \tag{5.13}$$

Because the correlation matrix is modeled to be positive definite, the objective function $\psi(\lambda)$ reaches its unique minimum value at the root of the partial derivative (5.12). Therefore, optimal weights for the Kriging predictor (5.6) can be derived by solving

$$R\lambda(x) + F\frac{\mu}{2\sigma^2} = r(x). \tag{5.14}$$

Together with the unbiasedness condition, this can be reformulated in matrix form

$$\begin{pmatrix} R & F \\ F^{\mathrm{T}} & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \tilde{\mu} \end{pmatrix} = \begin{pmatrix} r(x) \\ f^{\mathrm{T}}(x) \end{pmatrix}, \tag{5.15}$$

where $\tilde{\mu} = \frac{\mu}{2\sigma^2}$ and $0 \in \mathbb{R}^{p \times p}$ denotes the zero matrix.

Since the correlation matrix R is invertible, the Schur complement can be used to invert the block matrix on the left hand side of Eq. (5.15), [31, p. 433],

$$\begin{pmatrix} R & F \\ F^{\mathrm{T}} & 0 \end{pmatrix}^{-1} = \begin{pmatrix} R^{-1} - R^{-1}F(F^{\mathrm{T}}R^{-1}F)^{-1}F^{\mathrm{T}}R^{-1} & R^{-1}F(F^{\mathrm{T}}R^{-1}F)^{-1} \\ (F^{\mathrm{T}}R^{-1}F)^{-1}F^{\mathrm{T}}R^{-1} & -(F^{\mathrm{T}}R^{-1}F)^{-1} \end{pmatrix}. \tag{5.16}$$

The Kriging predictor can be rewritten in terms of Eq. (5.16),

$$\hat{y}(x) = \lambda^{\mathrm{T}}(x)Y = \begin{pmatrix} r^{\mathrm{T}}(x) & f(x) \end{pmatrix} \begin{pmatrix} R & F \\ F^{\mathrm{T}} & 0 \end{pmatrix}^{-1} \begin{pmatrix} Y \\ 0 \end{pmatrix}. \tag{5.17}$$

Inserting Eq. (5.16) into Eq. (5.17) yields

$$\hat{y}(x) = f(x)\hat{\beta} + r^{\mathrm{T}}(x)\,\underbrace{R^{-1}(Y - F\hat{\beta})}_{=:v}, \tag{5.18}$$

a simpler form of the Kriging predictor, where

$$\hat{\beta} := \left(F^{\mathrm{T}}R^{-1}F\right)^{-1} F^{\mathrm{T}}R^{-1}Y, \tag{5.19}$$

and

$$v = R^{-1}(Y - F\hat{\beta}) \tag{5.20}$$

are independent of the location x. Thus, the linear systems (5.19) and (5.20) need to be solved only once. After that, the evaluation of the predictor (5.18) at the location x takes the effort of one evaluation of the regression function $f(x)$, n evaluations of the correlation kernel $\mathcal{R}_\theta(x, \hat{x})$ to determine the correlation vector $r(x)$ and two Euclidean inner products of size p and n.

Note that $\hat{\beta}$ in (5.19) is the common generalized least squares estimate for the regression parameter β, [70, p. 413].

5.1.3. Maximum likelihood estimation of the Kriging parameters

As mentioned above, the choice of the correlation parameter θ has a large impact on the Kriging predictor. Therefore it is usually determined by a maximum likelihood prediction, cf. Definition 2.5. Following the assumptions that the error $\varepsilon(x)$ is a Gaussian process and we observed a realization of the random vector $(\varepsilon(x^1), \dots, \varepsilon(x^n))^{\mathrm{T}}$, the likelihood function is

$$L(\theta, \sigma^2, \beta) = \frac{1}{\sqrt{\det(R) \cdot (2\pi\sigma^2)^n}} \exp\left(-\frac{(Y - F\beta)^{\mathrm{T}} R^{-1} (Y - F\beta)}{2\sigma^2} \right), \tag{5.21}$$

cf. Sec. 2.1, [45, p. 163] and [28, p. 55]. By applying the natural logarithm, we obtain the simpler expression

$$\ln(L)(\theta, \sigma^2, \beta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2}\ln(\det(R)) - \frac{(Y - F\beta)^{\mathrm{T}} R^{-1} (Y - F\beta)}{2\sigma^2}. \tag{5.22}$$

Since the natural logarithm is strictly increasing, the maxima of $\ln(L)$ and L occur at the same locations. For a fixed correlation parameter θ, we can therefore obtain likelihood-optimal Kriging parameters σ^2 and β by setting the partial derivatives of $\ln(L)$ to zero. This yields

$$\hat{\beta} = \left(F^{\mathrm{T}} R^{-1} F\right)^{-1} F^{\mathrm{T}} R^{-1} Y \tag{5.23}$$

and

$$\hat{\sigma}^2 = \frac{(Y - F\hat{\beta})^{\mathrm{T}} R^{-1} (Y - F\hat{\beta})}{n}. \tag{5.24}$$

Under the assumption that the correlation matrix is positive definite, the parameter $\hat{\sigma}^2$ in Eq. (5.24) is positive whenever the observed responses Y can not be interpolated by any regression model of the chosen type $f(x)\beta$. In this case, the corresponding Hessian is negative definite at $(\hat{\beta}, \hat{\sigma}^2)$, and thus these parameters maximize the likelihood function L for a fixed correlation parameter θ. On the other hand, if it is possible to interpolate the observed responses by a regression model $f(x)\beta$, there is simply no need of fitting a Kriging model. So this is not a severe restriction.

Note that the likelihood-optimal regression parameter $\hat{\beta}$ in Eq. (5.23) equals the one we have set in Eq. (5.19) and thus the generalized least squares estimate for β.

Substituting $\hat{\beta}$ and $\hat{\sigma}^2$ back into Eq. (5.22) and removing constant terms, we derive the so called *concentrated ln-likelihood function*, [28, p. 55],

$$\ln(L) \approx -\frac{n}{2}\ln(\hat{\sigma}^2) - \ln(\det(R)), \tag{5.25}$$

which can now be used to find likelihood-optimal correlation parameters θ. The optimization is done numerically, since the derivatives of (5.25) with respect to θ do not have a simple structure. For numerical optimization, Forrester et al. suggested a genetic algorithm or simulated annealing, [28, p. 55], while Lophaven et al. used a modification of a direct search method proposed by Hooke and Jeeves for their MATLAB Kriging Toolbox, [52, pp. 29]. We found that in our applications, the numerical optimization of the correlation parameters is challenging, because of extremly different behavior of the target function in different regions of the search space.

Remark 5.1. *Inserting the likelihood-optimal parameters $\hat{\sigma}^2$ and $\hat{\beta}$, as well as the Kriging weights $\lambda(x)$ found in Eq. (5.17) into Eq. (5.8), one can derive a predictor for the mean squared error of the Kriging estimator at a location $x \in \mathcal{D}$,*

$$\mathrm{MSE}[x] = \hat{\sigma}^2 \left(1 + u^{\mathrm{T}}(F^{\mathrm{T}}R^{-1}F)^{-1}u - r^{\mathrm{T}}(x)R^{-1}r(x)\right), \tag{5.26}$$

where $u = F^{\mathrm{T}}R^{-1}r(x) - f(x)$, [53, p. 4].

Note that this estimator predicts the mean squared error based on the estimated spatial correlation of the input variables only. The estimated mean squared error is small at locations close to sample points and grows with the distance to sample sites. Its maximum value is $\hat{\sigma}^2$, [48, p. 6]. For that reason it is capable of detecting gaps in the sampling but it usually does not provide quantitatively accurate error information.

5.2. Cokriging

This section is based on Section 2–4 of our article [6].

As with any other interpolation method, the construction of a Kriging model requires sampled data of the original model which is to approximate. However, an evaluation of the original model is often highly computationally expensive and thus it is sometimes not feasible to compute sufficiently many sample values to obtain a surrogate model of adequate accuracy. This high computational effort can be significantly reduced, if in addition to the original high-accurate model a second model is available, which is less accurate but cheaper to evaluate. The idea of so-called *variable-fidelity* modeling techniques is to use the cheaper model to obtain a global trend of the quantity of interest in the design space and use this trend to interpolate the high-accurate sampled data. Models of different levels of accuracy which describe the same quantity often exist in applications. Think for example of two different ways to compute the drag coefficient introduced in Sec. 3.3 for a test body via flow simulations: One by performing a RANS simulation and one by using the much more expensive but also more accurate DES, cf. Sec. 3.2. Another example is the modeling of scalar quantities which are obtained by solving a partial differential equation (PDE) discretized on a fine spatial grid as opposed to solving the same equation on a coarser grid.

Approaches which deal with sampled data of different levels of accuracy go also under the name *multi-fidelity* surrogate modeling in literature, [23]. An extension of the Kriging method to variable-fidelity modeling is known as *Cokriging*. A variety of different Cokriging approaches has been developed by the geostatistical community, [40], and the statistical community, [42, 64]. The approach of [42] has been successfully applied to computer experiments in the context of aerospace design in [27]. In order to develop a simple Cokriging method for the use in aerodynamic analysis and optimization, [36, 37, 84] introduced an alternative version of Cokriging, which has a simple notation and follows closely the formulation of Kriging in the context of the design and analysis of computer experiments, [28, 70, 71].

After a recapitulation of the Cokriging approach introduced in [36], open theoretical problems associated with this approach are addressed: The positive definiteness of the Cokriging covariance matrix is investigated in Section 5.2.3. In Section 5.2.4, analytical expressions for likelihood-optimal Cokriging parameters are derived and an optimality condition is given.

5.2.1. Basic model assumptions of Cokriging

Suppose, we aim to approximate a function $y_1 \colon \mathcal{D} \to \mathbb{R}$ on the design variable space $\mathcal{D} \subset \mathbb{R}^d$, the so-called *high-fidelity model*, which is highly accurate but expensive to evaluate. In addition, the function $y_2\colon \mathcal{D} \to \mathbb{R}$, hereinafter referred to as the *low-fidelity model*, describes the same quantity as y_1, is cheaper to evaluate but less accurate. Similar to the basic assumption of Kriging, cf. Sec. 5.1.1, it is assumed that the true high- and low-fidelity functions y_1 and y_2 are realizations of dependent random functions

$$\begin{aligned} \mathcal{Y}_1(x) &= f_1(x)\beta_1 + \varepsilon_1(x), \\ \mathcal{Y}_2(x) &= f_2(x)\beta_2 + \varepsilon_2(x), \end{aligned} \tag{5.27}$$

where $f_1(x), f_2(x)\colon \mathbb{R}^d \to \mathbb{R}^p$ are regression models with corresponding regression parameter vectors $\beta_1, \beta_2 \in \mathbb{R}^p$ and $\varepsilon_1(x), \varepsilon_2(x)$ are Gaussian processes with expectation values $\mathrm{E}[\varepsilon_1(x)] = \mathrm{E}[\varepsilon_2(x)] = 0$ and stationary variances $\mathrm{Var}[\varepsilon_1(x)] = \sigma_1^2 > 0$ and $\mathrm{Var}[\varepsilon_2(x)] = \sigma_2^2 > 0$. Furthermore, it is presumed that the auto- and cross-covariances of the Gaussian processes $\varepsilon_1(x)$ and $\varepsilon_2(x)$ are given by a known positive definite correlation kernel $\mathcal{R}_\theta(x, \tilde{x})$, so that

$$\begin{aligned} \mathrm{Cov}[\varepsilon_1(x), \varepsilon_1(\tilde{x})] &= \sigma_1^2 \mathcal{R}_\theta(x, \tilde{x}), \\ \mathrm{Cov}[\varepsilon_1(x), \varepsilon_2(\tilde{x})] &= \sigma_1 \sigma_2 \mathcal{R}_\theta(x, \tilde{x}), \\ \mathrm{Cov}[\varepsilon_2(x), \varepsilon_2(\tilde{x})] &= \sigma_2^2 \mathcal{R}_\theta(x, \tilde{x}), \end{aligned} \tag{5.28}$$

with unknown correlation parameters $\theta \in \mathbb{R}_+^d$.

5.2.2. Construction of the Cokriging predictor

The high-fidelity model y_1 and the low-fidelity model y_2 are evaluated at a set of sample locations $\mathcal{X} = \{x^1, \dots, x^{n_1}\} \subset \mathbb{R}^d$ and $\tilde{\mathcal{X}} = \{\tilde{x}^1, \dots, \tilde{x}^{n_2}\} \subset \mathbb{R}^d$, respectively (usually $n_2 \gg n_1$). Let us first assume that low- and high-fidelity locations are pairwise distinct, $\mathcal{X} \cap \tilde{\mathcal{X}} = \emptyset$. Later in this work we will see how we can avoid this restriction. The corresponding high- and low-fidelity function values are given by

$$\begin{aligned} Y_1 &= [y^1, \dots, y^{n_1}]^{\mathrm{T}} := [y_1(x^1), \dots, y_1(x^{n_1})]^{\mathrm{T}} \in \mathbb{R}^{n_1}, \\ Y_2 &= [\tilde{y}^1, \dots, \tilde{y}^{n_2}]^{\mathrm{T}} := [y_2(\tilde{x}^1), \dots, y_2(\tilde{x}^{n_2})]^{\mathrm{T}} \in \mathbb{R}^{n_2}. \end{aligned} \tag{5.29}$$

Let $F_1 \in \mathbb{R}^{n_1 \times p}$ and $F_2 \in \mathbb{R}^{n_2 \times p}$ be the matrices of regression function values of f_1 and f_2 at the sample locations $\mathcal{X}$ and $\tilde{\mathcal{X}}$, respectively. Furthermore, let $\Sigma_1 \in \mathbb{R}^{n_1}$ and

$\Sigma_2 \in \mathbb{R}^{n_2}$ be the corresponding errors, such that

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} F_1 & 0 \\ 0 & F_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \Sigma_1 \\ \Sigma_2 \end{pmatrix} =: F\beta + \Sigma. \tag{5.30}$$

The objective of Cokriging is to interpolate the data of high-fidelity by using the trend of the low-fidelity data and taking spatial correlation into account. As Kriging, Cokriging yields the BLUE: The Cokriging estimator is constructed to be linear in the observed data Y_1 and Y_2,

$$\hat{y}(x) = \lambda_1(x)^{\mathrm{T}} Y_1 + \lambda_2(x)^{\mathrm{T}} Y_2 = \begin{pmatrix} \lambda_1(x)^{\mathrm{T}}, & \lambda_2(x)^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} =: \lambda(x)^{\mathrm{T}} Y \tag{5.31}$$

with weights $\lambda_1(x) \in \mathbb{R}^{n_1}$ and $\lambda_2(x) \in \mathbb{R}^{n_2}$ computed such that the estimator minimizes the mean squared error $\mathrm{E}[(y_1(x) - \hat{y}(x))^2]$, while being unbiased with respect to the high-fidelity process, i.e., $\mathrm{E}[y_1(x) - \hat{y}(x)] = 0$. By using Eq. (5.30) and the linearity of the expected value, the unbiasedness condition reads

$$\begin{aligned} 0 &= \mathrm{E}[y_1(x) - \hat{y}(x)] \\ &= \lambda_1(x)^{\mathrm{T}} F_1 \beta_1 + \lambda_2(x)^{\mathrm{T}} F_2 \beta_2 + \lambda_1(x)^{\mathrm{T}} \underbrace{\mathrm{E}[\Sigma_1]}_{=0} + \lambda_2(x)^{\mathrm{T}} \underbrace{\mathrm{E}[\Sigma_2]}_{=0} - f_1(x)\beta_1 - \underbrace{\mathrm{E}[\varepsilon_1(x)]}_{=0} \end{aligned}$$

and therefore

$$(\lambda_1(x)^{\mathrm{T}} F_1 - f_1(x))\beta_1 + \lambda_2(x)^{\mathrm{T}} F_2 \beta_2 = 0.$$

In order to fulfill the unbiasedness constraint independent of the choice of the regression parameters β_1 and β_2, the following stronger unbiasedness condition must be imposed:

$$\lambda_1^{\mathrm{T}} F_1 - f_1(x) = 0, \qquad \lambda_2^{\mathrm{T}} F_2 = 0, \tag{5.32}$$

where $\lambda_i \equiv \lambda_i(x), i = 1, 2$. Due to the linearity of the expected value and by using the unbiasedness constraints (5.32), the mean squared error becomes

$$\begin{aligned}
\mathrm{MSE}(x) &= \mathrm{E}\left[(\hat{y}(x) - y_1(x))^2\right] \\
&= (\lambda^{\mathrm{T}} F\beta)^2 - 2\lambda^{\mathrm{T}} F\beta f_1(x)\beta_1 + \lambda^{\mathrm{T}} \mathrm{E}\left[\Sigma\Sigma^{\mathrm{T}}\right]\lambda - 2\lambda^{\mathrm{T}} \mathrm{E}\left[\Sigma\varepsilon_1(x)\right] \\
&\qquad + (f_1(x)\beta_1)^2 + \mathrm{E}\left[\varepsilon_1(x)^2\right] \\
&= \lambda^{\mathrm{T}} F\beta(\lambda^{\mathrm{T}} F\beta - f_1(x)\beta_1) + f_1(x)\beta_1(f_1(x)\beta_1 - \lambda^{\mathrm{T}} F\beta) + \lambda^{\mathrm{T}} C\lambda \\
&\qquad - 2\lambda^{\mathrm{T}} c(x) + \sigma_1^2 \\
&= \lambda^{\mathrm{T}} C\lambda - 2\lambda^{\mathrm{T}} c(x) + \sigma_1^2,
\end{aligned} \tag{5.33}$$

where $\lambda \equiv \lambda(x) = (\lambda_1^{\mathrm{T}}(x), \lambda_2^{\mathrm{T}}(x))^{\mathrm{T}}$ and

$$C := \begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix}$$

is the auto- and cross-covariance matrix, which, according to our choice of the correlation kernel (5.28), reads

$$\begin{aligned}
C^{11} &= \sigma_1^2 R^{11} := \sigma_1^2 \left(\mathcal{R}_\theta(x^i, x^j)\right)_{i,j=1,\dots,n_1} \in \mathbb{R}^{n_1 \times n_1}, \\
C^{12} &= \sigma_1\sigma_2 R^{12} := \sigma_1\sigma_2 \left(\mathcal{R}_\theta(x^i, \tilde{x}^j)\right)_{i=1,\dots,n_1, j=1,\dots,n_2} \in \mathbb{R}^{n_1 \times n_2} \\
C^{21} &= \sigma_1\sigma_2 R^{12} = \left(R^{12}\right)^{\mathrm{T}} \in \mathbb{R}^{n_2 \times n_1}, \\
C^{22} &= \sigma_2^2 R^{22} := \sigma_2^2 \left(\mathcal{R}_\theta(\tilde{x}^i, \tilde{x}^j)\right)_{i,j=1,\dots,n_2} \in \mathbb{R}^{n_2 \times n_2}.
\end{aligned}$$

Moreover, the quantity $c(x)$ in (5.33) is

$$c(x) := \begin{pmatrix} c_1(x) \\ c_2(x) \end{pmatrix},$$

which is the vector of covariances of a given point x and the sample points. More precisely,

$$\begin{aligned} c_1(x) &= \sigma_1^2 r_1(x) := \sigma_1^2 \left(\mathcal{R}_\theta(x^i, x) \right)_{i=1,\dots,n_1} \in \mathbb{R}^{n_1}, \\ c_2(x) &= \sigma_1\sigma_2 r_2(x) := \sigma_1\sigma_2 \left(\mathcal{R}_\theta(\tilde{x}^i, x) \right)_{i=1,\dots,n_2} \in \mathbb{R}^{n_2}. \end{aligned}$$

To account for the unbiasedness constraints (5.32), Lagrange multipliers $\mu_1, \mu_2 \in \mathbb{R}^p$ are introduced. The associated MSE minimization problem reads

$$\min_\lambda \lambda^\mathrm{T} C\lambda - 2\lambda^\mathrm{T} c(x) + \sigma_1^2 + \mu_1 \left(\lambda_1^\mathrm{T} F_1 - f_1(x) \right) + \mu_2 \lambda_2^\mathrm{T} F_2. \tag{5.34}$$

The first-order optimality condition is given by the following system of linear equations:

$$\begin{bmatrix} C^{11} & C^{12} & F_1 & 0 \\ C^{12} & C^{22} & 0 & F_2 \\ F_1^\mathrm{T} & 0^\mathrm{T} & 0 & 0 \\ 0^\mathrm{T} & F_2^\mathrm{T} & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \frac{1}{2}\mu_1 \\ \frac{1}{2}\mu_2 \end{bmatrix} = \begin{bmatrix} c_1(x) \\ c_2(x) \\ f_1^\mathrm{T}(x) \\ 0 \end{bmatrix}. \tag{5.35}$$

The solution of this linear system is the unique solution of the global minimization problem (5.34), if the corresponding Hessian is positive definite and the matrix on the left hand side of Eq. (5.35) has full rank. We will see in Section 5.2.3 that these conditions are fulfilled only in the case of pairwise distinct high- and low-fidelity locations. If this assumption does not hold, modifications need to be done.

In order to rewrite the system in terms of correlations rather than covariances, Han et al. [36] introduced a parameter transformation,

$$\tilde{\lambda}_1 := \lambda_1, \qquad \tilde{\lambda}_2 := \frac{\sigma_2}{\sigma_1}\lambda_2, \qquad \tilde{\mu}_1 := \frac{1}{2\sigma_1^2}\mu_1, \qquad \tilde{\mu}_2 := \frac{1}{2\sigma_1\sigma_2}\mu_2. \tag{5.36}$$

This leads to an equivalent system of equations,

$$\begin{bmatrix} R^{11} & R^{12} & F_1 & 0 \\ R^{12} & R^{22} & 0 & F_2 \\ F_1^\mathrm{T} & 0^\mathrm{T} & 0 & 0 \\ 0^\mathrm{T} & F_2^\mathrm{T} & 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{\lambda}_1 \\ \tilde{\lambda}_2 \\ \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{bmatrix} = \begin{bmatrix} r_1(x) \\ r_2(x) \\ f_1^\mathrm{T}(x) \\ 0 \end{bmatrix}. \tag{5.37}$$

Solving this system of linear equations by using the Schur complement inversion, see Lemma 2.18, yields

$$\hat{y}(x) = f_1(x)\tilde{\beta}_1 + r^{\mathrm{T}}(x) \underbrace{R^{-1}(\tilde{y}_S - F\tilde{\beta})}_{=:v_{cok}}, \tag{5.38}$$

where

$$\begin{aligned} \tilde{\beta} = \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{bmatrix} = \left(F^{\mathrm{T}} R^{-1} F\right)^{-1} F^{\mathrm{T}} R^{-1} \tilde{y}_S, \quad r(x) = \begin{bmatrix} r_1(x) \\ r_2(x) \end{bmatrix}, \\ R = \begin{bmatrix} R^{11} & R^{12} \\ R^{21} & R^{22} \end{bmatrix}, \quad \tilde{y}_S = \begin{bmatrix} y_1 \\ \frac{\sigma_1}{\sigma_2} y_2 \end{bmatrix}. \end{aligned} \tag{5.39}$$

Since the vector v_{cok} in (5.38) does not depend on the spatial location x, a Cokriging prediction requires only the evaluation of two inner products.

Remark 5.2. *As with Kriging, one can derive an estimator for the mean squared error from Eq. (5.33) by using the substitutions in Eq. (5.36) and the expression for the Cokriging weights $\lambda(x)$ derived by the Schur complement inversion:*

$$\mathrm{MSE}[x] = \sigma_1^2 \left(1 + u^{\mathrm{T}}(F^{\mathrm{T}} R^{-1} F)^{-1} u - r^{\mathrm{T}}(x) R^{-1} r(x)\right), \tag{5.40}$$

where $u = F^{\mathrm{T}} R^{-1} r(x) - [f_1(x), 0]^{\mathrm{T}}$, cf. [34, p. 7] for the case of a constant regression function.

An essential requirement for the well-posedness of the Cokriging predictor (5.38) is the positive definiteness of the auto- and cross-correlation matrix R. This issue will be addressed in the next subsection.

5.2.3. On the positve definiteness of the Cokriging correlation matrix

This Section is based on our articles [72] and [6, Sec. 3].

A unique solution of the system of Cokriging equations (5.35) corresponds to a global solution of the MSE minimization problem (5.34), if this optimization problem is convex, that is, if the corresponding Hessian is positive definite. Obviously, (up to a constant factor), the Hessian is precisely the covariance matrix C. In this section, necessary conditions to obtain a well-defined Cokriging predictor are derived.

Our first lemma relates the positive definiteness of the correlation matrix R with the positive definiteness of the covariance matrix C.

Lemma 5.3. *Let the auto- and cross correlation matrix R and the corresponding covariance matrix C be defined as above, so that*

$$R = \begin{pmatrix} R^{11} & R^{12} \\ R^{21} & R^{22} \end{pmatrix}, \quad \begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 R^{11} & \sigma_1\sigma_2 R^{12} \\ \sigma_1\sigma_2 R^{21} & \sigma_2^2 R^{22} \end{pmatrix}.$$

Then the matrices R and C are congruent. In particular, the covariance matrix C is positive definite, if and only if the correlation matrix R is positive definite.

Proof. The following decomposition shows that C is congruent to the matrix $\sigma_1^2 R$:

$$\begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix} = \begin{pmatrix} I_{n_1} & 0 \\ 0 & \frac{\sigma_2}{\sigma_1} I_{n_2} \end{pmatrix}^T \cdot \sigma_1^2 \begin{pmatrix} R^{11} & R^{12} \\ R^{21} & R^{22} \end{pmatrix} \cdot \begin{pmatrix} I_{n_1} & 0 \\ 0 & \frac{\sigma_2}{\sigma_1} I_{n_2} \end{pmatrix}$$

As a result, due to Sylvester's law of inertia, the covariance matrix C is positive definite, if and only if the correlation matrix R is positive definite. □

Suppose that the correlation is modeled via a strictly positive definite correlation kernel $\mathcal{R}_\theta(x, \hat{x})$ with the same hyper-parameter θ for the auto- and cross-correlation model. Then, the correlation matrix is guaranteed to be positive definite provided that the low- and high-fidelity sample points are mutually distinct, $\mathcal{X} \cap \tilde{\mathcal{X}} = \emptyset$, cf. Lemma 2.7.[7] As an immediate consequence, Lemma 5.3 implies the following:

Corollary 5.4. *Let the low- and high-fidelity sample locations be pairwise distinct, e.g. $\mathcal{X} \cap \tilde{\mathcal{X}} = \emptyset$ and suppose that the auto- and cross-covariance is modeled via the same strictly positive definite correlation kernel. Then, the correlation matrix R and the covariance matrix C are positive definite. Thus the Cokriging predictor is well defined and optimal in the sense that it minimizes the mean squared error.*

There are, however, many VFM application scenarios, where low-fidelity data is readily available at the high-fidelity sample locations so that $\mathcal{X} \cap \tilde{\mathcal{X}} \neq \emptyset$ and the above result does not apply. For example, consider applications where the same iterative solver is used to compute both the high- and the low-fidelity samples and the low-fidelity values correspond to a relaxed numerical convergence threshold.[8] Hence, lower-fidelity data is inherently available at every high-fidelity sample point

---

[7] If different values for the hyper-parameter θ are used to model the auto- and cross-correlations, Lemma 2.7 does not apply. In practice, one may still obtain a positive definite correlation matrix. However, this is not guaranteed.

[8] This is, e.g., the setting considered in [18].

at no extra computational costs. Such double sample points give valuable additional information on the functional relation between the high- and low-fidelity model. There are VFM techniques such as additive or multiplicative bridge functions [35], which even require that $\mathcal{X} \subset \tilde{\mathcal{X}}$. However, when the double sample locations are input in the Cokriging correlation matrix, the matrix will feature equal columns and will thus be rank-deficient. In [36, § III.B, p. 10] it was observed by means of numerical experiments that this problem is avoided when the cross-correlation blocks R^{12} and R^{21} of the correlation matrix are scaled down by a factor of $\rho \in (0, 1)$. Below, a proof that such a scaling indeed yields a positive definite correlation matrix is given.

Theorem 5.5. *Let $\mathcal{X} = \{x^1, \dots, x^{n_1}\}$ and $\tilde{\mathcal{X}} = \{\tilde{x}^1, \dots, \tilde{x}^{n_2}\}$ be arbitrary high- and low-fidelity sample points with $x^i \neq x^j$ and $\tilde{x}^i \neq \tilde{x}^j$ provided that $i \neq j$. Furthermore, let $\mathcal{R}_\theta(x, \hat{x})$ be a strictly strictly positive definite correlation kernel. Then, for every $\rho \in [0, 1)$, the correlation matrix*

$$R_\rho = \begin{pmatrix} R^{11} & \rho R^{12} \\ \rho R^{21} & R^{22} \end{pmatrix}$$

is positive definite.

Proof. Here, a proof for the case $\mathcal{X} \subset \tilde{\mathcal{X}}$ is given. The more general case of $\tilde{\mathcal{X}} \cap \mathcal{X} \neq \emptyset$ is technically more complex but follows exactly the same ideas. The proof for the case $\tilde{\mathcal{X}} \cap \mathcal{X} = \emptyset$ is a part of the proof for $\mathcal{X} \subset \tilde{\mathcal{X}}$.

W.l.o.g., we assume that the coincident high-fidelity and low-fidelity points are the last ones in the low-fidelity sample set, i.e., $x^1 = \tilde{x}^{n_2 - n_1 + 1}, \dots, x^{n_1} = \tilde{x}^{n_2}$. Let $\mathcal{Z} := \{\tilde{x}^1, \dots, \tilde{x}^{n_2 - n_1}\}$, so that $\mathcal{Z} \cup \mathcal{X} = \tilde{\mathcal{X}}$ and $\mathcal{Z} \cap \mathcal{X} = \emptyset$. Furthermore, let

$$\begin{aligned} R^{xx} &:= R^{11} = (\mathcal{R}_\theta(x^i, x^j))_{i,j=1,\dots,n_1}, \\ R^{xz} &= (R^{zx})^\mathrm{T} = (\mathcal{R}_\theta(x^i, \tilde{x}^j))_{i=1,\dots,n_1, j=1,\dots,n_2-n_1}, \\ R^{zz} &= (\mathcal{R}_\theta(\tilde{x}^i, \tilde{x}^j))_{i,j=1,\dots,n_2-n_1}. \end{aligned}$$

With this notation we have

$$R_\rho = \left(\begin{array}{c|cc} R^{xx} & \rho R^{xz} & \rho R^{xx} \\ \hline \rho R^{zx} & R^{zz} & R^{zx} \\ \rho R^{xx} & R^{xz} & R^{xx} \end{array}\right) = \left(\begin{array}{cc|c} R^{xx} & \rho R^{xz} & \rho R^{xx} \\ \rho R^{zx} & R^{zz} & R^{zx} \\ \hline \rho R^{xx} & R^{xz} & R^{xx} \end{array}\right) =: \begin{pmatrix} \tilde{R}^{11} & \tilde{R}^{12} \\ \tilde{R}^{21} & \tilde{R}^{22} \end{pmatrix}.$$

Due to the Schur complement condition for positive definite matrices, cf. Lemma 2.19 and [31, Prop. 16.1, p. 434], the matrix R_ρ is positive definite if and only if the matrix $\tilde{R}^{22}$ and the Schur complement $\hat{R} := \tilde{R}^{11} - \tilde{R}^{12}(\tilde{R}^{22})^{-1}\tilde{R}^{21}$ are positive definite. Because $\mathcal{R}_\theta(x,\hat{x})$ is a strictly positive definite correlation kernel and the elements of $\mathcal{X}$ are pairwise distinct, the matrix $\tilde{R}^{22} = R^{xx}$ is positive definite. Moreover,

$$\begin{aligned}\hat{R} &= \begin{pmatrix} R^{xx} & \rho R^{xz} \\ \rho R^{zx} & R^{zz} \end{pmatrix} - \begin{pmatrix} \rho R^{xx} \\ R^{zx} \end{pmatrix} (R^{xx})^{-1} \begin{pmatrix} \rho R^{xx} & R^{xz} \end{pmatrix} \\ &= \begin{pmatrix} (1-\rho^2)R^{xx} & 0 \\ 0 & R^{zz} - R^{zx}(R^{xx})^{-1}R^{xz} \end{pmatrix}.\end{aligned}$$

The upper block $(1-\rho^2)R^{xx}$ is clearly positive definite, whenever $0 \leq \rho < 1$. The lower block $R^{zz} - R^{zx}(R^{xx})^{-1}R^{xz}$ is a Schur complement of the matrix block R^{22}. This block, in turn, is positive definite because it is the auto-correlation matrix of the pairwise distinct low-fidelity points and $\mathcal{R}_\theta(x,\hat{x})$ is a strictly positive definite kernel. Therefore $\hat{R}$ is positive definite and so is the correlation matrix R_ρ. □

Remark 5.6. *The above Theorem formally includes the special case of $\rho = 0$. As a matter of fact, the Cokriging predictor reduces to the standard Kriging predictor for $\rho = 0$. This will be discussed in detail in the upcoming Corollary 5.12.*

5.2.4. Maximum likelihood estimation of the Cokriging parameters

This section is based on our article [6, Sec. 4].

The Cokriging predictor (5.38) depends on a number of model- and hyper-parameters, namely the process variances σ_1, σ_2, the vectors $\tilde{\beta}_1, \tilde{\beta}_2$ of regression coefficients and the range parameter vector θ appearing in the spatial correlation model (5.28). These parameters are determined by the method of maximum likelihood estimation [28, 36, 71].

The likelihood function is given by

$$L(\tilde{\beta}, \sigma_1/\sigma_2, \sigma_1^2, \theta) = \frac{1}{\sqrt{(2\pi\sigma_1^2)^{n_1+n_2}\,|R|}} \exp\left(-\frac{1}{2\sigma_1^2}\left(\tilde{y}_S - F\tilde{\beta}\right)^{\mathrm{T}} R^{-1}\left(\tilde{y}_S - F\tilde{\beta}\right)\right) \tag{5.41}$$

(see [36, § II.D, p. 7 et sqq.] for a derivation). Optimal values for $\tilde{\beta}, \sigma_1/\sigma_2$ and σ_1^2 can be derived via optimization along profile lines, i.e., via optimizing with respect to a single parameter (vector) while fixing the remaining parameters, see, e.g., [28, p. 169].

This yields

$$\tilde{\beta} = \begin{bmatrix}\tilde{\beta}_1 \\ \tilde{\beta}_2\end{bmatrix} = \left(F^{\mathrm{T}} R^{-1} F\right)^{-1} F^{\mathrm{T}} R^{-1} \begin{bmatrix} y_1 \\ \frac{\sigma_1}{\sigma_2} y_2 \end{bmatrix}, \tag{5.42}$$

$$\frac{\sigma_1}{\sigma_2} = -\left(\begin{bmatrix}0 \\ y_2\end{bmatrix}^{\mathrm{T}} R^{-1} \begin{bmatrix}0 \\ y_2\end{bmatrix}\right)^{-1} \begin{bmatrix}0 \\ y_2\end{bmatrix}^{\mathrm{T}} R^{-1} \left(\begin{bmatrix}y_1 \\ 0\end{bmatrix} - F\tilde{\beta}\right), \tag{5.43}$$

$$\sigma_1^2 = \frac{1}{n_1 + n_2} \cdot \left(\begin{bmatrix} y_1 \\ \frac{\sigma_1}{\sigma_2} y_2 \end{bmatrix} - F\tilde{\beta}\right)^{\mathrm{T}} R^{-1} \left(\begin{bmatrix} y_1 \\ \frac{\sigma_1}{\sigma_2} y_2 \end{bmatrix} - F\tilde{\beta}\right). \tag{5.44}$$

Observe that the likelihood-optimal values for $\tilde{\beta}$ and σ_1/σ_2 still feature a mutual dependence. In [36, § III, p. 11], a numerical fixed point iteration was used to optimize the parameters $\tilde{\beta}, \sigma_1/\sigma_2$ and σ_1^2, taking the mutual dependence and the dependence on the range parameter θ into account.

Here, closed form solutions for likelihood-optimal parameters $\tilde{\beta}, \sigma_1/\sigma_2$ and σ_1^2 are presented that depend only on the correlation length parameter θ. Hence, three model parameters can be dropped from the numerical likelihood optimization process.

In order to resolve the mutual dependency, Eq. (5.42) is substituted into Eq. (5.43). By setting $R_F := R^{-1}F$ and $M := F^{\mathrm{T}} R^{-1} F$ this yields

$$\begin{aligned}
\frac{\sigma_1}{\sigma_2} &= \left(\begin{bmatrix}0 \\ y_2\end{bmatrix}^{\mathrm{T}} R^{-1} \begin{bmatrix}0 \\ y_2\end{bmatrix}\right)^{-1} \begin{bmatrix}0 \\ y_2\end{bmatrix}^{\mathrm{T}} R^{-1} \left(\begin{bmatrix}-y_1 \\ 0\end{bmatrix} + F\left(M^{-1} R_F^{\mathrm{T}} \begin{bmatrix} y_1 \\ \frac{\sigma_1}{\sigma_2} y_2 \end{bmatrix}\right)\right) \\
&= \left(\begin{bmatrix}0 \\ y_2\end{bmatrix}^{\mathrm{T}} R^{-1} \begin{bmatrix}0 \\ y_2\end{bmatrix}\right)^{-1} \begin{bmatrix}0 \\ y_2\end{bmatrix}^{\mathrm{T}} R^{-1} \begin{bmatrix}-y_1 \\ 0\end{bmatrix} \\
&\quad + \left(\begin{bmatrix}0 \\ y_2\end{bmatrix}^{\mathrm{T}} R^{-1} \begin{bmatrix}0 \\ y_2\end{bmatrix}\right)^{-1} \begin{bmatrix}0 \\ y_2\end{bmatrix}^{\mathrm{T}} R_F M^{-1} R_F^{\mathrm{T}} \begin{bmatrix}y_1 \\ 0\end{bmatrix} \\
&\quad + \frac{\sigma_1}{\sigma_2} \left(\begin{bmatrix}0 \\ y_2\end{bmatrix}^{\mathrm{T}} R^{-1} \begin{bmatrix}0 \\ y_2\end{bmatrix}\right)^{-1} \begin{bmatrix}0 \\ y_2\end{bmatrix}^{\mathrm{T}} R_F M^{-1} R_F^{\mathrm{T}} \begin{bmatrix}0 \\ y_2\end{bmatrix}.
\end{aligned} \tag{5.45}$$

Subtracting the last term on both sides of the equation leads to

$$\begin{aligned}
&\underbrace{\frac{\sigma_1}{\sigma_2}\left(1-\left(\begin{bmatrix}0\\y_2\end{bmatrix}^{\mathrm{T}} R^{-1}\begin{bmatrix}0\\y_2\end{bmatrix}\right)^{-1}\begin{bmatrix}0\\y_2\end{bmatrix}^{\mathrm{T}} R_F M^{-1} R_F^{\mathrm{T}}\begin{bmatrix}0\\y_2\end{bmatrix}\right)}_{:=z}\\
&\qquad=\left(\begin{bmatrix}0\\y_2\end{bmatrix}^{\mathrm{T}} R^{-1}\begin{bmatrix}0\\y_2\end{bmatrix}\right)^{-1}\begin{bmatrix}0\\y_2\end{bmatrix}^{\mathrm{T}}\left(R_F M^{-1} R_F^{\mathrm{T}}-R^{-1}\right)\begin{bmatrix}y_1\\0\end{bmatrix}. \qquad (5.46)
\end{aligned}$$

If $z \neq 0$, division by z results in

$$\frac{\sigma_1}{\sigma_2}=\frac{\begin{bmatrix}0\\y_2\end{bmatrix}^{\mathrm{T}}\left(R_F M^{-1} R_F^{\mathrm{T}}-R^{-1}\right)\begin{bmatrix}y_1\\0\end{bmatrix}}{\begin{bmatrix}0\\y_2\end{bmatrix}^{\mathrm{T}}\left(R^{-1}-R_F M^{-1} R_F^{\mathrm{T}}\right)\begin{bmatrix}0\\y_2\end{bmatrix}}. \qquad (5.47)$$

This expression only depends on the correlation parameter θ. After determining σ_1/σ_2, one can compute the likelihood-optimal values for $\tilde{\beta}$ and σ_1^2 according to (5.42) and (5.44), respectively. The following theorem ensures the well-posedness of this procedure.

Theorem 5.7. *Suppose the correlation matrix R is positive definite and that the matrix*

$$\hat{F} := \left[F, \begin{bmatrix}y_1\\0\end{bmatrix}, \begin{bmatrix}0\\y_2\end{bmatrix}\right] \qquad (5.48)$$

has full column rank. Then the parameter $\varrho^* := \sigma_1/\sigma_2$ in (5.47) is well-defined. Furthermore let

$$\tilde{\beta}^* := \tilde{\beta}\left(\varrho^*\right) = \left(F^{\mathrm{T}} R^{-1} F\right)^{-1} F^{\mathrm{T}} R^{-1}\begin{bmatrix}y_1\\ \varrho^* y_2\end{bmatrix}, \qquad (5.49)$$

$$\left(\sigma_1^2\right)^* := \sigma_1^2\left(\varrho^*, \tilde{\beta}^*\right) = \frac{1}{n_1+n_2}\cdot\left(\begin{bmatrix}y_1\\ \varrho^* y_2\end{bmatrix} - F\tilde{\beta}^*\right)^{\mathrm{T}} R^{-1}\left(\begin{bmatrix}y_1\\ \varrho^* y_2\end{bmatrix} - F\tilde{\beta}^*\right). \qquad (5.50)$$

Then, for every fixed choice of θ, the parameters $\nu^* := \left(\varrho^*, \tilde{\beta}^*, \left(\sigma_1^2\right)^*\right)$ are global maximizers of the likelihood function (5.41).

Proof. The denominator of σ_1/σ_2 in (5.47) is a Schur complement of the matrix

$$\begin{aligned}\tilde{H} &= \left[\begin{array}{c|c} F^{\mathrm{T}}R^{-1}F & -F^{\mathrm{T}}R^{-1}\begin{bmatrix}0\\ y_2\end{bmatrix} \\ \hline -\begin{bmatrix}0\\ y_2\end{bmatrix}^{\mathrm{T}} R^{-1}F & \begin{bmatrix}0\\ y_2\end{bmatrix}^{\mathrm{T}} R^{-1}\begin{bmatrix}0\\ y_2\end{bmatrix}\end{array}\right] \\ &= \begin{bmatrix} F^{\mathrm{T}} \\ -\begin{bmatrix}0\\ y_2\end{bmatrix}^{\mathrm{T}}\end{bmatrix} R^{-1}\left[F,\ -\begin{bmatrix}0\\ y_2\end{bmatrix}\right] =: \hat{F}_1^{\mathrm{T}}R^{-1}\hat{F}_1, \end{aligned} \tag{5.51}$$

which is positive definite because the extended matrix $\hat{F}$ has full column rank and thus the nullspace of $\hat{F}_1$ is trivial.

Due to the Schur complement condition for the positive definiteness of block matrices [31, Prop. 16.1], the Schur complement of the matrix $\tilde{H}$ must also be positive definite and therefore the denominator of σ_1/σ_2 in (5.47) is strictly positive and ϱ^* is well-defined.

It is clear from the construction of ϱ^*, $\tilde{\beta}^*$ and $\left(\sigma_1^2\right)^*$ that the first-order partial derivatives of the likelihood function (5.41) with respect to σ_1/σ_2, $\tilde{\beta}$ and σ_1^2 vanish at ν^*. To guarantee the optimality at ν^*, we need to show that the Hessian matrix H is negative definite. The Hessian matrix of the likelihood function (5.41) with respect to σ_1/σ_2, $\tilde{\beta}$ and σ_1^2 is

$$\begin{aligned} H &= \underbrace{-\frac{\exp\left(-\frac{1}{2}(n_1+n_2)\right)}{\sqrt{(2\pi\sigma_1^2)^{n_1+n_2}|R|\cdot\sigma_1^2}}}_{=:c}\cdot\left[\begin{array}{c|c|c} F^{\mathrm{T}}R^{-1}F & -F^{\mathrm{T}}R^{-1}\begin{bmatrix}0\\ y_2\end{bmatrix} & 0 \\ \hline -\begin{bmatrix}0\\ y_2\end{bmatrix}^{\mathrm{T}} R^{-1}F & \begin{bmatrix}0\\ y_2\end{bmatrix}^{\mathrm{T}} R^{-1}\begin{bmatrix}0\\ y_2\end{bmatrix} & 0 \\ \hline 0^{\mathrm{T}} & 0 & \frac{n_1+n_2}{2\sigma_1^2}\end{array}\right] \\ &= c\cdot\left[\begin{array}{c|c} \tilde{H} & 0 \\ \hline 0 & \frac{n_1+n_2}{2\sigma_1^2}\end{array}\right], \end{aligned} \tag{5.52}$$

with $\tilde{H}$ as introduced in (5.51). Because $\hat{F}$ has full column rank, for every regression parameter vector β, there is at least one high-fidelity sample value $y^i \in Y_1$ which is not interpolated exactly by the regression model $f_1(x)\beta$ alone. Therefore, the vector $y_s - F\tilde{\beta}$ is non-zero so that σ_1^2 in (5.44) is strictly positive. Hence,

$$c = -\frac{\exp\left(-\frac{1}{2}(n_1+n_2)\right)}{\sqrt{(2\pi\sigma_1^2)^{n_1+n_2}|R|}\cdot\sigma_1^2}$$

is well defined and negative. Because of $\frac{n_1+n_2}{2\sigma_1^2} > 0$ and the positive definiteness of the matrix $\tilde{H}$ (see above), the Hessian matrix H is negative definite. Thus, the parameters ν^* indeed maximize the likelihood function (5.41). □

In practical applications, the hyper-parameter θ of the correlation model $\mathcal{R}_\theta(x,\tilde{x})$ is determined by a numerical maximum likelihood estimation [36, p. 9]. Instead of optimizing the likelihood function (5.41) directly, the so-called *condensed log-likelihood function* [62, p. 675]

$$\ln(L) = -\frac{1}{2}\left[(n_1+n_2)\cdot\ln(\sigma_1^2) + \ln\det(R)\right] \tag{5.53}$$

is considered, which is obtained from the likelihood function (5.41) by applying the natural logarithm and neglecting constant terms, [36, p. 9]. Obviously, the likelihood function and the condensed log-likelihood function share the same locations for their local and global optima. Moreover, in [36, 84] it was proposed to treat the cross-correlation scaling factor ρ also as a statistical quantity that is to be included in the likelihood optimization process. According to the exposition in Section 5.2.3, we suggest that ρ should rather be considered as a numerical relaxation parameter that is to be chosen close to 1 but so that the correlation matrix does not degenerate numerically. A numerical examination of this issue can be found in Section 5.2.6.

Because there is no analytic solution for the optimal hyper-parameter θ, it is determined by a numerical optimization of the condensed log-likelihood function (5.53). However, Theorem 5.7 allows us to compute optimal values for $\sigma_1/\sigma_2, \sigma_1^2$ and $\tilde{\beta}$ depending on the correlation parameters θ and ρ in closed form at each iteration step of the likelihood optimization.

Remark 5.8. *The requirement that* $\hat{F} = \left[F, \begin{bmatrix} y_1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ y_2 \end{bmatrix}\right]$ *has full column rank means that neither the high-fidelity nor the low-fidelity sample values can be interpolated exactly by the chosen regression model alone, i.e., without considering the Gaussian error term in (5.38). This is a natural assumption, because otherwise, it is not necessary to fit a Cokriging model in the* *first place. In practice, it is often not known a priori, if the quantities of interest features a linear, polynomial or general non-linear parameter dependency. However, even if the parameter dependency is highly nonlinear, this assumption yields a restriction on the choice of the regression model as the following example illustrates.*

Consider a two-parameter problem, e. g. we aim to optimize the lift-to-drag ration of a wing profile depending on its height and length. When a linear regression model $f_1(x)\beta = \beta^0 + \beta^1 x_1 + \beta^2 x_2$ *is employed, any given three high-fidelity sample points* (x^i, y_i), $x^i \in \mathbb{R}^2$, $i = 1, 2, 3$, *are interpolated exactly with the following choice of regression coefficients:*

$$\underbrace{\begin{bmatrix} 1 & x_1^1 & x_2^1 \\ 1 & x_1^2 & x_2^2 \\ 1 & x_1^3 & x_2^3 \end{bmatrix}}_{=F_1} \begin{bmatrix} \beta^0 \\ \beta^1 \\ \beta^2 \end{bmatrix} = \underbrace{\begin{bmatrix} y(x^1) \\ y(x^2) \\ y(x^3) \end{bmatrix}}_{=y_1} .$$

An example for such a scenario is displayed in Figure 6.

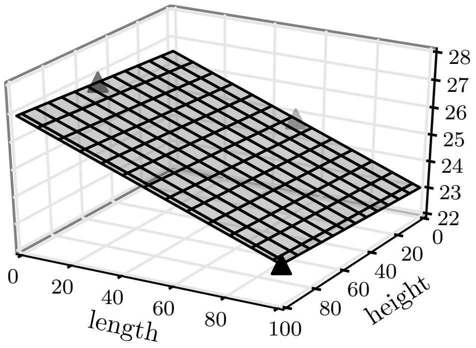

Figure 6.: Three sample points in a two dimensional parameter space can exactly be interpolated by a linear regression model.

In this situation it does not matter if the quantity of interest actually features a linear dependency on the input variables or if the global trend implied by the low-fidelity data is highly nonlinear. The high-fidelity sample points can be interpolated exactly by a linear regression model which implies that (F_1, y_1) *and thus* $\hat{F}$ *is rank-deficient. As a result, Theorem 5.7 does not apply. The consequences of the rank-deficiency on the analytical expressions (5.47)–(5.50) are discussed in the following. In the special case of this example, a reduction of the complexity (the degrees of freedom) of the regression model or the computation of additional high-fidelity sample points may overcome the problem of rank-deficiency.*

In the following, the different cases of rank-deficiency of the matrix $\hat{F}$ in (5.48) are addressed.

Observation 5.9. *Suppose the correlation matrix R is positive definite and the matrix $[F_2, y_2]$ is rank-deficient, i.e. the low-fidelity sample points can be described by a regression model of the chosen type. Then the analytical expression (5.47) for the Cokriging parameter σ_1/σ_2 is undefined.*

Proof. Suppose the low-fidelity sample data can be interpolated exactly by a regression model of the chosen type $f_2(x)$. Then there exists a vector of regression coefficients $\gamma \in \mathbb{R}^p$ such that $F_2\gamma = y_2$ or

$$\begin{bmatrix} 0 \\ y_2 \end{bmatrix} = \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix} \underbrace{\begin{bmatrix} 0 \\ \gamma \end{bmatrix}}_{=:\tilde{\gamma}} = F\tilde{\gamma}.$$

Substitution in the denominator of σ_1/σ_2 in (5.47) yields,

$$\begin{aligned} &\begin{bmatrix} 0 \\ y_2 \end{bmatrix}^{\mathrm{T}} \left(R^{-1} - R^{-1}F(F^{\mathrm{T}}R^{-1}F)^{-1}F^{\mathrm{T}}R^{-1}\right) \begin{bmatrix} 0 \\ y_2 \end{bmatrix} \\ &\qquad = \begin{bmatrix} 0 \\ y_2 \end{bmatrix}^{\mathrm{T}} (R^{-1}F\tilde{\gamma} - R^{-1}F \underbrace{(F^{\mathrm{T}}R^{-1}F)^{-1}F^{\mathrm{T}}R^{-1}F}_{=I_p} \tilde{\gamma}) \\ &\qquad = 0. \end{aligned}$$

Thus, the analytical expression (5.47) is undefined. □

Observation 5.10. *Let the correlation matrix R be positive definite. Assume that $[F_2, y_2]$ has full column rank but $[F_1, y_1]$ is rank-deficient, i.e. the observed high-fidelity data can be described exactly by a regression model $f_1(x)\alpha$. Then, the analytical expressions for the Cokriging parameters (5.47)–(5.50) are well-defined, although they do not represent a valid solution to the maximum likelihood problem. If these parameters are used to fit a Cokriging model, the corresponding predictor takes the form*

$$\hat{y}(x) = f_1(x)\alpha. \tag{5.54}$$

Proof. Suppose the high-fidelity sample data can be interpolated exactly by a regression model of the chosen type $f_1(x)$. Then there exists a vector of regression coefficients $\alpha \in \mathbb{R}^p$ such that $F_1\alpha = y_1$ or

$$\begin{bmatrix} y_1 \\ 0 \end{bmatrix} = \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix} \underbrace{\begin{bmatrix} \alpha \\ 0 \end{bmatrix}}_{=:\tilde{\alpha}} = F\tilde{\alpha}.$$

Because $[F_2, y_2]$ has full column rank, the matrix $\hat{F}_1$ in (5.51) is also not rank-deficient and thus the analytical expressions (5.47)–(5.50) for the Cokriging parameters are well-defined (but not guaranteed to be likelihood-optimal). Substitution in the numerator of σ_1/σ_2 in (5.47) yields,

$$\begin{aligned} &\begin{bmatrix} 0 \\ y_2 \end{bmatrix}^{\mathrm{T}} \left(R^{-1}F(F^{\mathrm{T}}R^{-1}F)^{-1}F^{\mathrm{T}}R^{-1} - R^{-1} \right) \begin{bmatrix} y_1 \\ 0 \end{bmatrix} \\ &\qquad = \begin{bmatrix} 0 \\ y_2 \end{bmatrix}^{\mathrm{T}} (R^{-1}F \underbrace{(F^{\mathrm{T}}R^{-1}F)^{-1}F^{\mathrm{T}}R^{-1}F}_{=I_p} \tilde{\alpha} - R^{-1}F\tilde{\alpha}) \\ &\qquad = 0. \end{aligned} \tag{5.55}$$

Thus, $\sigma_1/\sigma_2 = 0$, which mathematically implies $\sigma_1^2 = 0$ and $\varepsilon_1(x) \equiv 0$. Notwithstanding this, the blind evaluation of the analytic expressions (5.49) and (5.50) yields

$$\tilde{\beta} = (F^{\mathrm{T}}R^{-1}F)^{-1}F^{\mathrm{T}}R^{-1} \begin{bmatrix} y_1 \\ 0 \cdot y_2 \end{bmatrix} = (F^{\mathrm{T}}R^{-1}F)^{-1}F^{\mathrm{T}}R^{-1}F\tilde{\alpha} = \tilde{\alpha} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix},$$

and

$$\sigma_1^2 = \frac{1}{n_1+n_2}(y_s - F\tilde{\beta})^{\mathrm{T}}R^{-1}(y_s - F\tilde{\beta}) = \frac{1}{n_1+n_2} 0^{\mathrm{T}} \cdot R^{-1} \cdot 0 = 0. \tag{5.56}$$

This is not a valid solution to the maximum likelihood problem since the likelihood function (5.41) is not even defined in the case $\sigma_1^2 = 0$. However, the analytical expressions can be evaluated. If this case is not being intercepted, the construction of a Cokriging predictor (5.38) yields

$$\hat{y}(x) = f_1(x) \underbrace{\tilde{\beta}_1}_{=\alpha} + r^{\mathrm{T}}(x)R^{-1}\underbrace{(\tilde{y}_S - F\tilde{\beta})}_{=0} = f_1(x)\alpha.$$

□

In general, the computation of the numerical rank of a matrix is computationally expensive compared to the other computations which are needed to set up a Cokriging predictor, see e.g. [32, § 5.4.1, § 5.4.3]. The conditions of the observations are thus hard to check in practice. The next observation provides a more useful access to the case described in Observation 5.10.

Observation 5.11. *Suppose that the correlation matrix R is positive definite and the matrix $[F_2, y_2]$ has full column rank. Then, the analytical expression (5.47) yields $\sigma_1/\sigma_2 = 0$, if and only if the high-fidelity sample points can be interpolated by a regression model of the chosen type.*

Proof. In the case that the high-fidelity sample points can be interpolated by a regression model of the chosen type, (5.56) in the proof of Observation 5.10 shows $\sigma_1/\sigma_2 = 0$. On the other hand, suppose that the analytical expression (5.55) yields $\sigma_1/\sigma_2 = 0$. This implies $\sigma_1^2 = 0$, see Eq. (5.56) in the proof of Observation 5.10, i.e.

$$\sigma_1^2 = \frac{1}{n_1 + n_2}(y_s - F\tilde{\beta})^{\mathrm{T}} R^{-1}(y_s - F\tilde{\beta}) = 0.$$

Due to the positive definiteness of the correlation matrix R, this yields

$$0 = (y_s - F\tilde{\beta}) = \begin{bmatrix} y_1 - F_1\tilde{\beta}_1 \\ \frac{\sigma_1}{\sigma_2} y_2 - F_2\tilde{\beta}_2 \end{bmatrix}$$

and thus $y_1 = F_1\tilde{\beta}_1$, which states that the observed high-fidelity sample points can be interpolated by a regression model of the chosen type. ☐

These observations can be used to determine whether the Cokriging method in combination with the selected regression model is generally applicable to a specific interpolation problem. If the denominator in the analytical expression (5.47) for the likelihood-optimal Cokriging parameter σ_1/σ_2 is determined to be 0, the low-fidelity sample points can be interpolated by the chosen regression model and do therefore not contain more information than the regression model itself. In that case, a single-fidelity Kriging model can be used to interpolate the high-fidelity data by considering the same regression model.

When the numerator in (5.47) is found to be 0, the high-fidelity sample points can be interpolated by the chosen regression model. Especially if the evaluation of the original high-fidelity model is computationally expensive, this may be caused by too few high-fidelity sample points in relation to the degrees of freedom of the selected regression model (as is the case in the example in Remark 5.8). The evaluation of the analytical expressions (5.47)–(5.50) is nevertheless possible, however the corresponding Cokriging predictor matches the interpolating regression model and thus does neither invoke any low-fidelity information nor take spatial correlations into account.

In all other cases, the analytical expressions (5.47)–(5.50) yield valid solutions to the maximum likelihood problem.

As announced in Section 5.2.3, Theorem 5.7 allows to cover the special case if the cross-correlation scaling factor is set to $\rho = 0$ in a corollary:

Corollary 5.12. *Suppose that the assumptions of Theorem 5.7 apply. Then the likelihood-optimal Cokriging predictor for the cross-correlation scaling factor $\rho = 0$ reduces precisely to the corresponding Kriging predictor that is based on the same regression function $f_1(x)$ and the same correlation model $\mathcal{R}_\theta(x, \hat{x})$.*

Proof. In the case $\rho = 0$, the correlation matrix R and thus its inverse R^{-1} reduce to block diagonal form,

$$R^{-1} = \begin{bmatrix} (R^{11})^{-1} & 0 \\ 0 & (R^{22})^{-1} \end{bmatrix}.$$

Because the matrix of regression values F features also block diagonal structure, the same holds true for the matrix

$$A := \left[R_F M^{-1} R_F^{\mathrm{T}} - R^{-1}\right] = \left[R^{-1}F(F^{\mathrm{T}}R^{-1}F)^{-1}F^{\mathrm{T}}R^{-1} - R^{-1}\right],$$

which occurs in the analytic expression for the likelihood-optimal Cokriging parameter σ_1/σ_2 in (5.47). Substitution of the block diagonal matrix A in (5.47) yields

$$\frac{\sigma_1}{\sigma_2} = \frac{\begin{bmatrix} 0 \\ y_2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} A^{11} & 0 \\ 0 & A^{22} \end{bmatrix} \begin{bmatrix} y_1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 0 \\ y_2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} -A^{11} & 0 \\ 0 & -A^{22} \end{bmatrix} \begin{bmatrix} 0 \\ y_2 \end{bmatrix}} = -\frac{0}{y_2^{\mathrm{T}} A^{22} y_2} = 0.$$

Note that the denominator $y_2^T A^{22} y_2$ is non-zero, because we assumed that the assumptions of Theorem 5.7 apply and thus the expression is well-defined. Because of the block diagonal structure of F and R^{-1}, the likelihood-optimal Cokriging parameter β in (5.49) reduces to

$$\tilde{\beta}^* = (F^{\mathrm{T}} R^{-1} F)^{-1} F^{\mathrm{T}} R^{-1} \begin{bmatrix} y_1 \\ 0 \cdot y_2 \end{bmatrix} = \begin{bmatrix} (F_1^{\mathrm{T}} (R^{11})^{-1} F_1)^{-1} F_1^{\mathrm{T}} (R^{11})^{-1} y_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \tilde{\beta}_1 \\ 0 \end{bmatrix}.$$

Substitution in the expression for the Cokriging estimator (5.38) yields,

$$\begin{aligned} \hat{y}(x) &= f_1(x)\tilde{\beta}_1 + \left[r_1(x)^{\mathrm{T}}, r_2(x)^{\mathrm{T}} \right] \begin{bmatrix} (R^{11})^{-1} & 0 \\ 0 & (R^{22})^{-1} \end{bmatrix} \left(\begin{bmatrix} y_1 \\ 0 \end{bmatrix} - \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix} \begin{bmatrix} \tilde{\beta}_1 \\ 0 \end{bmatrix} \right), \\ &= f_1(x)\tilde{\beta}_1 + r_1(x)^{\mathrm{T}} (R^{11})^{-1} \left(y_1 - F_1 \tilde{\beta}_1 \right), \end{aligned} \tag{5.57}$$

which does not depend on the low-fidelity data. Furthermore, if the same regression model $f(x)$ and the same correlation kernel $\mathcal{R}_\theta(x, \hat{x})$ (including the same correlation parameter θ) is taken to set up a Kriging predictor based on the high-fidelity data $Y \equiv y_1$, the corresponding likelihood-optimal Kriging parameter $\hat{\beta}$ in (5.23) coincides with the Cokriging parameter $\tilde{\beta}_1$. Thus, the Kriging estimator (5.6) equals exactly the Cokriging estimator (5.57) in this case. □

These observations clarify that the Cokriging predictor (5.38) indeed generalizes Kriging and regression.

5.2.5. Numerical validation of the findings

This subsection is based on Section 5.2 and 5.3 of our article [6].

For the purpose of validation of the findings presented in the previous subsections, the academic test case from [28, p. 173] is considered. It is demonstrated that the closed form likelihood optima derived in Section 5.2.3 indeed coincide with their numerically optimized counterparts obtained by the fixed point iteration of [36, § III, p. 11]. Furthermore, it is shown that including low-fidelity samples also at locations where high-fidelity samples are available, i.e., $\mathcal{X} \subset \tilde{\mathcal{X}}$ is beneficial.

The academic example established itself as a standard test case and has been studied in e.g. [18,36]. However, it should be mentioned that this example exhibits a special linear relation between the associated high- and low-fidelity models that is hardly ever encountered in real-life applications. Hence, the almost perfect performance of many VFM models for this example may be deceiving and cannot be expected to transfer to real-life applications, see also [18].

parameter	analyt.	fixed point it.	difference (Eucl. norm)
σ_1/σ_2	2.0000	2.0000	$6.2222 \cdot 10^{-10}$
σ_1^2	31.3683	31.3682	$1.3733 \cdot 10^{-4}$
$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$	$\begin{bmatrix} 0.6262 \\ 5.9453 \\ -19.3738 \\ 25.9453 \end{bmatrix}$	$\begin{bmatrix} 0.6262 \\ 5.9453 \\ -19.3738 \\ 25.9453 \end{bmatrix}$	$6.2281 \cdot 10^{-5}$

Table 2.: Resulting Cokriging parameters determined using the two different approaches and the Euclidean norm of their respective differences

The high-fidelity function reads

$$y_1\colon [0,1] \to \mathbb{R}, x \mapsto (6x-2)^2 \sin(12x-4), \tag{5.58}$$

the associated low-fidelity function is set to

$$y_2\colon [0,1] \to \mathbb{R}, x \mapsto 0.5\, y_1(x) + 10\,(x-0.5) + 5. \tag{5.59}$$

The same sampling as in [36, § V.A, p. 13 et sqq.] with $n_1 = 3$ high-fidelity and $n_2 = 11$ low-fidelity sample points is considered: three high-fidelity samples were computed at $\mathcal{X} = \{0.0, 0.6, 1.0\}$ and eleven low-fidelity samples were computed at equidistant locations $\tilde{\mathcal{X}} = \{0.0, 0.1, \ldots, 0.9, 1.0\}$. Note that every high-fidelity sample location also appears as a low-fidelity sample location, i.e., $\mathcal{X} \subset \tilde{\mathcal{X}}$.

To demonstrate that the analytically determined Cokriging parameters coincide with those obtained by the fixed point iteration introduced in [36, § III, p. 11], two Cokriging models were fitted based on the given data sets. For the first model, the Cokriging parameters were computed using (5.47)–(5.50). For the second model, the fixed point iteration was used. In both cases, a linear regression model $f(x)\beta = \beta^0 + \beta^1 x$ and the Gaussian correlation kernel

$$\mathcal{R}_\theta(x, \hat{x}) = e^{-\theta|x-\hat{x}|^2} \tag{5.60}$$

were employed and the cross-correlation scaling factor was set to $\rho = 1.0 - 10^{-10}$.

The analytically determined Cokriging parameters and those determined by the fixed point iteration introduced in [36, § III, p. 11] are listed in Table 2.

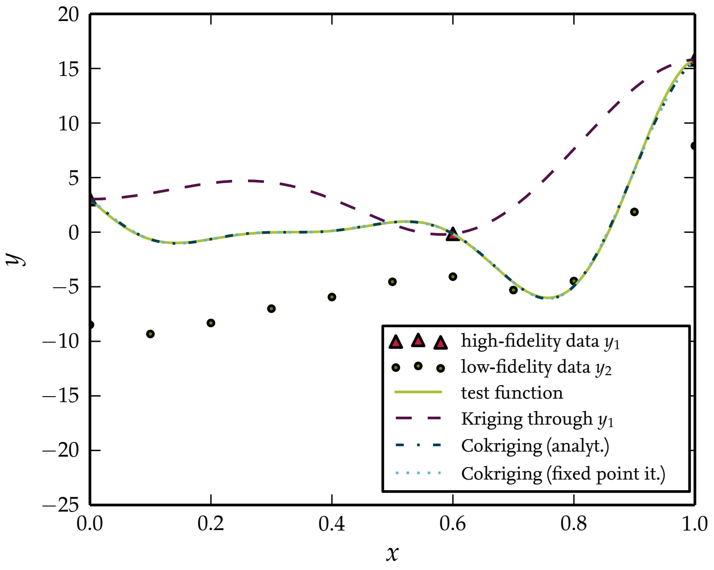

Figure 7.: Approximation of the analytical test function using Kriging and Cokriging. For the latter, the predictor associated with the closed-form optimal MLE parameters is compared with the predictor where these parameters were found via a numerical fixed point iteration.

As one can see from Table 2, the Cokriging parameters coincide up to a numerical accuracy of order 10^{-4}. The likelihood-optimal correlation parameter θ was found to be $\theta = 23.6364$.[9] A plot of the resulting Cokriging predictors is shown in Figure 7. For comparison, a Kriging model was constructed using only the three high-fidelity sample points, which is also displayed in Figure 7.

As both the numerically optimized and the closed-form parameters virtually coincide, so do the corresponding Cokriging predictors.

Next, the benefits of an inclusive low-fidelity sampling $\mathcal{X} \subset \tilde{\mathcal{X}}$ are demonstrated. To this end, the same three high-fidelity sample points of $\mathcal{X}$ were kept but the low-fidelity samples were reduced to $\tilde{\mathcal{X}}_2 = \tilde{\mathcal{X}} \setminus \mathcal{X}$ with $n_{2,2} = 8$ sample locations. Although the correlation matrix R is nonsingular even for $\rho = 1$, the cross-correlation scaling factor ρ was set to $1.0 - 10^{-10}$ for better comparison to the previous case. Figure 8 shows the resulting Cokriging predictor in comparison with the Cokiging predictor based on the inclusive sampling $\tilde{\mathcal{X}}$.

---

[9] The numerical optimization of the correlation parameter θ is difficult as the objective function is multimodal, numerically often unstable and since the search may get stuck in local optima. In order not to deviate from the main purpose of this work and to ensure that the results are not corrupted by insufficient numerical optima, the correlation parameters θ was determined by a *brute force exhaustive*

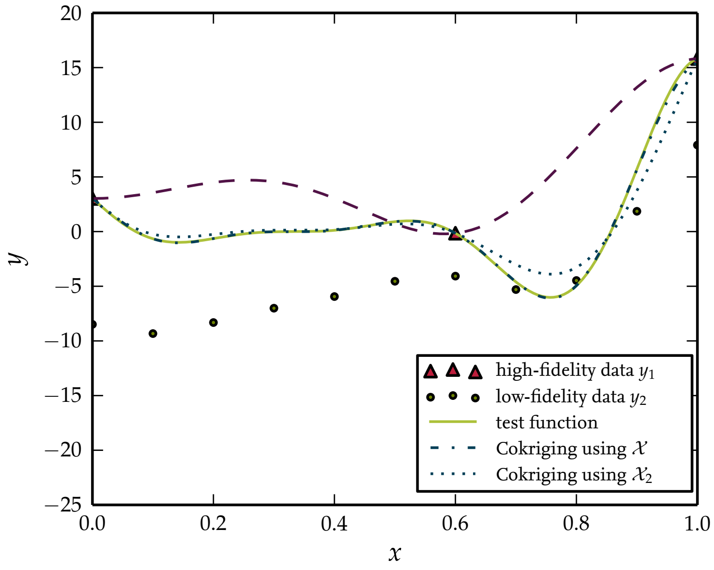

Figure 8.: Approximation of the analytical test function using Cokriging with two different low-fidelity data sets: the original one $\tilde{\mathcal{X}}$, used in the first test case, and the reduced low-fidelity sample set $\tilde{\mathcal{X}}_2 = \tilde{\mathcal{X}} \setminus \mathcal{X}$

As in the previous case, both Cokriging methods yield the same parameters and thus the same predictor. The correlation parameter θ was found to be 10.0000. Although the approximation given by the Cokriging predictors in the second case is still good, the results in the first test case are significantly better. For comparison, an approximation to the L^2 norm of the error

$$\|y(x) - \hat{y}(x)\|_{L^2} = \left(\int_0^1 (y(x) - \hat{y}(x))^2 \,\mathrm{d}x \right)^{1/2} \tag{5.61}$$

was computed by quadrature with the rectangle rule based on an equidistant sampling at 1000 sites $\hat{x}_i = 0.001 \cdot i, i = 0, \ldots, 999$,

$$\|y(x) - \hat{y}(x)\|_{L^2} \approx \left(\frac{1}{1000} \sum_{i=0}^{999} (y(\hat{x}_i) - \hat{y}(\hat{x}_i))^2 \right)^{1/2} . \tag{5.62}$$

By this procedure, the error of the first model using $\tilde{\mathcal{X}}$ as low-fidelity sample set was found to be $\|y(x) - \hat{y}(x)\|_{L^2} \approx 0.1092$, while the error of the second model using $\tilde{\mathcal{X}}_2$ as low-fidelity sample set is $\|y(x) - \hat{y}(x)\|_{L^2} \approx 1.0683$, which is 10 times

---

search on a full factorial grid of 100 sample sites per coordinate direction.

higher. Thus, the Cokriging predictor clearly benefits from an inclusive low-fidelity sampling.

5.2.6. Numerical investigation of the cross-correlation scaling factor ρ

To investigate the dependency of the Cokriging predictor on the cross-correlation scaling factor ρ, the analytical test case presented in the previous subsection is considered. While the correlation parameter θ is fixed to $\theta = 23.6364$ (which was found to be likelihood-optimal for $\rho = 1.0 - 10^{-10}$ in the previous subsection), the cross correlation scaling factor ρ was varied between $\rho = 0.0$ and $\rho = 1.0 - 10^{-7}$. Again, a linear regression model and a Gaussian correlation kernel were used. The Cokriging parameters were determined using the closed-form expressions (5.47)–(5.50). For comparison purposes, a Kriging predictor $\hat{y}_K(x)$ was built by using the same parameter settings, but only the high-fidelity sample points. In Figure 9, Cokriging predictors for $\rho \in \{0.0, 0.5, 0.9, 0.99, 0.999, 0.9999999\}$ and the Kriging model are displayed.

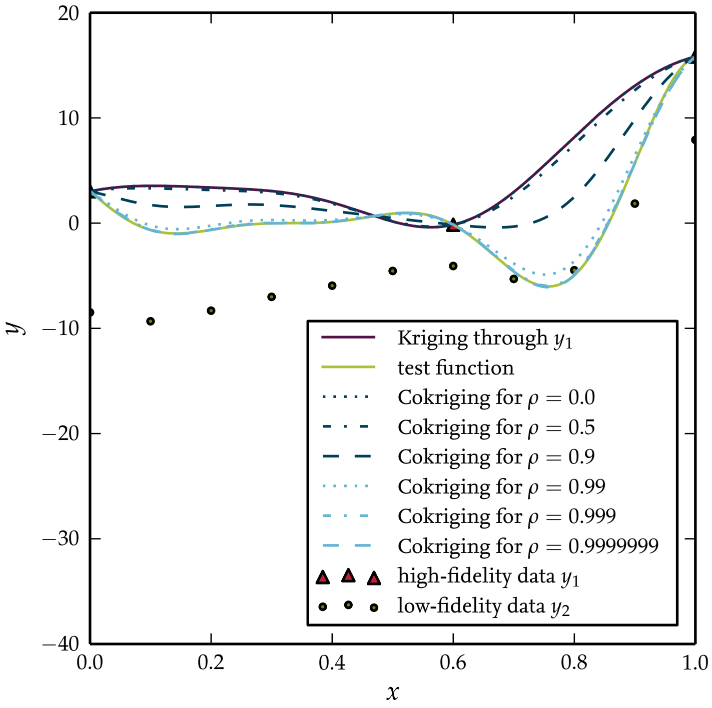

Figure 9.: Cokriging predictors for different choices of the cross-correlation scaling factor ρ

For better judging the experiments, approximations to the L_2 error between the different Cokriging models and the test function $y(x)$ were computed, employing the quadrature formula (5.62) as before. The Kriging model $\hat{y}_K(x)$ was also included in the comparison. The results are listed in Table 3. In addition, Figure 10 displays

ρ	approx. to $\|y(x) - \hat{y}(x)\|_{L^2}$	approx. to $\|\hat{y}_K(x) - \hat{y}(x)\|_{L^2}$
0.0	5.7526	$3.8853 \cdot 10^{-11}$
0.5	5.4273	0.3254
0.9	3.2566	2.4969
0.99	0.5924	5.1694
0.999	0.1256	5.6895
0.9999999	0.1092	5.7529

Table 3.: Approximation to the error $\|y(x) - \hat{y}_\rho(x)\|_{L^2}$ and the L_2 norm of the difference to the Kriging model for the Cokriging predictors for different values of the cross-correlation scaling factor ρ, using eq. (5.62)

the corresponding condensed log-likelihood function (5.53) as a function of the cross-correlation scaling factor ρ.

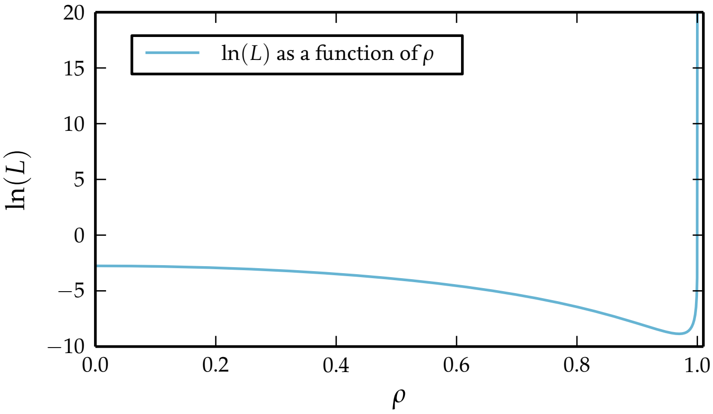

Figure 10.: Condensed log-likelihood function for the analytical test case with fixed correlation parameter $\theta = 23.6364$ as a function of the cross-correlation scaling factor ρ

As one can see from Figure 9 and the results in Table 3, the scaling factor ρ has a large impact on the Cokriging predictor. The higher the cross-correlation scaling factor is, the more is the trend of the low-fidelity sample points taken into account. In accordance with the theoretical result given in Corollary 5.12, the Cokriging predictor for $\rho = 0$ coincides up to a numerical accuracy of the order 10^{-11} with the Kriging predictor which does not consider any low-fidelity information. As the scaling factor ρ increases, the local extrema which are indicated by the course of the low-fidelity data become increasingly pronounced in the Cokriging predictor.

These results suggest that the scaling factor ρ can be seen as a regularization parameter which allows to control the impact of the low-fidelity sample points on the Cokriging predictor. However, Figure 10 shows that the condensed log-likelihood function, which, as mentioned above, is commonly used to determine the hyperparameter θ numerically, increases rapidly for $\rho \to 1$. Simultaneously, the correlation matrix R becomes more and more ill-conditioned and thus the numerical error introduced when solving linear systems with R grows. This can be taken as an additional indication that the cross-correlation scaling factor ρ is to be treated as a regularization parameter rather than a statistical quantity as its likelihood optimum seems to be at $\rho = 1$. Note also that in the case at hand, a numerical optimization of ρ will either move towards the lower bound $\rho = 0$ or the upper bound $\rho = 1$, depending on the chosen start location. This behavior seems to be typical as we have observed the same progression of the condensed log-likelihood function also in different other test cases, cf. [6, Fig. 8]. The parameter ρ should therefore not be included in the numerical likelihood optimization process.

5.3. Hierarchical Kriging

Another Kriging-based variable-fidelity surrogate modeling approach goes under the name Hierarchical Kriging, [34]. It is based on the simple idea of using an ordinary Kriging model for the low-fidelity data as regression model for the Kriging interpolation of the high-fidelity data.

As before, suppose the high-fidelity model y_1 and the low-fidelity model y_2 are sampled at a set of locations $\mathcal{X} = \{x^1, \ldots, x^{n_1}\} \subset \mathbb{R}^d$ and $\tilde{\mathcal{X}} = \{\tilde{x}^1, \ldots, \tilde{x}^{n_2}\} \subset \mathbb{R}^d$, respectively, the corresponding responses being

$$\begin{aligned} Y_1 &= [y^1, \ldots, y^{n_1}]^{\mathrm{T}} := [y_1(x^1), \ldots, y_1(x^{n_1})]^{\mathrm{T}} \in \mathbb{R}^{n_1}, \\ Y_2 &= [\tilde{y}^1, \ldots, \tilde{y}^{n_2}]^{\mathrm{T}} := [y_2(\tilde{x}^1), \ldots, y_2(\tilde{x}^{n_2})]^{\mathrm{T}} \in \mathbb{R}^{n_2}. \end{aligned} \tag{5.63}$$

As a first step, a Kriging model is constructed based on the low-fidelity data only. Following the approach described in Sec. 5.1, we obtain

$$\hat{y}_2(x) = f_2(x)\hat{\beta}_2 + r_2^{\mathrm{T}}(x) R_2^{-1}(Y_2 - F_2\hat{\beta}_2), \tag{5.64}$$

where $f_2\colon \mathbb{R}^d \to \mathbb{R}^p$ is a regression model with regression parameters $\hat{\beta}_2 \in \mathbb{R}^p$, $F_2 = (f_2(\tilde{x}^1), \dots, f_2(\tilde{x}^{n_2}))^{\mathrm{T}} \in \mathbb{R}^{n_2 \times p}$ is the matrix of regression function values, $R_2 \in \mathbb{R}^{n_2 \times n_2}$ is the correlation matrix of the low-fidelity sampled locations $\tilde{\mathcal{X}}$ and $r_2(x) \in \mathbb{R}^{n_2}$ is the vector of correlations between the location x and the sampled locations $\tilde{\mathcal{X}}$.

Under the assumption that the low-fidelity sample points give a good trend of the quantity of interest, this low-fidelity Kriging model is a suitable regression model for the Kriging interpolation of the high-accurate data. Following this idea, the resulting Hierarchical Kriging surrogate model for y_1 reads

$$\hat{y}(x) = \hat{y}_2(x)\hat{\beta} + r^{\mathrm{T}}(x)R^{-1}(Y_1 - F_1\hat{\beta}), \tag{5.65}$$

where $\hat{\beta} \in \mathbb{R}$ is a regression parameter which acts as a scaling factor for the low-fidelity model, $F_1 = (\hat{y}_2(x^1), \dots, \hat{y}_2(x^{n_1}))^{\mathrm{T}}$ is the matrix of regression function values, $R \in \mathbb{R}^{n_1 \times n_1}$ is the correlation matrix of the high-fidelity sample locations $\mathcal{X}$ and $r(x) \in \mathbb{R}^{n_1}$ is the vector of correlations between the location x and the sampled locations $\mathcal{X}$.

6. Variable-fidelity surrogate modeling of multi-dimensional responses

In this section, the variable-fidelity methodology is transferred to the setting of multi-dimensional responses.[10] A simple and well-known approach for the construction of non-intrusive single-fidelity surrogate models for high-dimensional outputs combines the low-rank approximation technique proper orthogonal decomposition (POD) and single-fidelity response surface models like Kriging, [9]. To invoke data of different levels of accuracy, we extended this approach to variable-fidelity response surfaces. After a short introduction to POD in Sec. 6.1, the new variable-fidelity surrogate modeling approach for vector-valued outputs is described in Sec. 6.2. Some technical remarks to this approach are given in Sec. 6.3, followed by a discussion on error estimation and the introduction of adaptive sampling strategies in Sec. 6.4.

6.1. Proper orthogonal decomposition

This subsection is based on [68, § 2.1] and [63, § 2].

Given a set of sampled data in a vector space $\mathcal{U}$, POD aims to find a k-dimensional subspace $\mathcal{U}_k \subset \mathcal{U}$ so that the error of projecting the sampled data onto the subspace $\mathcal{U}_k$ is minimized. Assume that the original model is of the form,

$$y\colon \mathcal{D} \times [0,T] \to \mathbb{R}^N, \tag{6.1}$$

with $N < \infty$ and $\mathcal{D} \subset \mathbb{R}^d$. For an introduction to POD in infinite-dimensional vector spaces, the reader is referred to [69].

The original model is sampled at a set of locations $\mathcal{X} = \{x^1, \dots, x^n\} \subset \mathcal{D}$ to obtain the trajectories $\mathcal{Y} := \{y_i\colon [0,T] \to \mathbb{R}^N, i = 1, \dots, n\}$. The goal is to find a projection $\Pi_k\colon \mathbb{R}^N \to \mathcal{U}_k$ of rank k that minimizes the total least-squares distance

$$\|\mathcal{Y} - \Pi_k \mathcal{Y}\|^2 := \sum_{i=1}^{n} \int_0^T \|y_i(t) - \Pi_k y_i(t)\|^2 \, \mathrm{d}t. \tag{6.2}$$

The solution to this problem is given by the following theorem, [63, Thm. 1]:

---

[10] In the remainder of this work, we will sometimes also use the term multiple-input multiple-output (MIMO) model for models with multi-dimensional responses, as common in the model order reduction community.

Theorem 6.1. *Let the symmetric, positive semi-definite matrix* $K \in \mathbb{R}^{N \times N}$ *be defined as,*

$$K := \sum_{i=1}^{n} \int_0^T y_i(t) y_i(t)^\mathrm{T} \, \mathrm{d}t, \tag{6.3}$$

and let the eigenvalues of K *be given in decending order,* $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. *Furthermore, let* $u^i \in \mathbb{R}^N, i = 1, \ldots, n$ *be the corresponding orthonormal eigenvectors, i.e.,*

$$K u^i = \lambda_i u^i, \tag{6.4}$$

and $\langle u^i, u^j \rangle = \delta_{ij}$ *for* $i, j = 1, \ldots, n$. *Then, the projection*

$$\Pi_k := \sum_{i=1}^{k} u^i (u^i)^\mathrm{T}, \tag{6.5}$$

minimizes the total least-squares distance $\|\mathcal{Y} - P\mathcal{Y}\|^2$ *in Eq. (6.2) among all projections* P *of rank* k *and it holds*

$$\|\mathcal{Y} - \Pi_k \mathcal{Y}\|^2 = \sum_{i=k+1}^{n} \lambda_i \tag{6.6}$$

Theorem 6.1 provides an instruction of how to determine a least-squares optimal k-dimensional subspace: The problem reduces to an eigenvalue problem. The orthonormal basis $\{u^1, \ldots, u^k\}$ of the subspace $\mathcal{U}_k$ is called *POD basis* and the vectors $u^i, i = 1, \ldots, k$ are called *POD modes*. The error formula in Eq. (6.6) is useful to answer the question of how to choose the dimension k of the approximating subspace: It motivates the definition of the *relative information content* of the first k POD modes, [63, Sec. 2.2, p. 98],

$$\mathrm{RIC}(k) := \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i}. \tag{6.7}$$

Usually, the POD basis is truncated after k POD modes, if the relative information content of these modes reaches a preset threshold, $\mathrm{RIC}(k) \geq 1 - \varepsilon$ with $0 < \varepsilon \leq 1$. It is clear that to obtain a good approximating and low-dimensional subspace $\mathcal{U}_k$ by means of a small least-squares error in Eq. (6.6) and a small dimension k, the eigenvalues of the matrix K need to decay sufficiently fast. In practice one indeed observes an exponential decay of the eigenvalues in many applications, [63, p. 99, Remark 2].

Remark 6.2. *Often one is interested in an affine subspace instead of a linear subspace, cf. [63, Remark 1, p. 98]. For example suppose we observe the flow fields of different car geometries* *in a wind tunnel. In huge areas of the wind tunnel, the flow fields will be quasi identical. In this case, it is therefore reasonable to view the overall flow field as a superposition of a mean flow and geometry dependent flow structures, which we are interested in. In this scenario, the matrix K in Eq. (6.3) is constructed in terms of the centered data instead of the original data, i.e. it is defined by*

$$K := \sum_{i=1}^{n} \int_0^T (y_i(t) - \overline{y})(y_i(t) - \overline{y})^{\mathrm{T}} \, \mathrm{d}t, \tag{6.8}$$

where $\overline{y} \in \mathbb{R}^N$ *is the mean value of the original data,*

$$\overline{y} := \frac{1}{nT} \sum_{i=1}^{n} \int_0^T y_i(t) \, \mathrm{d}t. \tag{6.9}$$

The affine linear space of dimension k *is then given by* $\overline{y} + \mathcal{U}_k$*, where* $\mathcal{U}_k$ *is the optimal* k*-dimensional subspace following the approach given in Theorem 6.1.*

The computation of the POD modes requires the solution of a $N \times N$ eigenvalue problem Eq. (6.4). When dealing with flow problems, the number of grid points and thus the dimension N is often very large and a direct solution of the eigenvalue problem is therefore not feasible.

A way out is proposed by Sirovich, [75], in the “method of snapshots”, [68, p. 4 et seq.]. If the data is given at discrete time instances $t_1, \ldots, t_m \in [0, T]$, the time integral in Eq. (6.3) is approximated by using a quadrature rule,

$$K \approx \tilde{K} = \sum_{i=1}^{n} \sum_{j=1}^{m} y^i(t_j) y^i(t_j)^{\mathrm{T}} \cdot \delta_j \tag{6.10}$$

with quadrature coefficients $\delta_1, \ldots, \delta_m > 0$. By setting

$$\tilde{Y} := \left[y^1(t_1)\sqrt{\delta_1}, \ \ldots, \ y^1(t_m)\sqrt{\delta_m}, \ \ \ldots, \ \ y^n(t_1)\sqrt{\delta_1}, \ \ldots, \ y^n(t_m)\sqrt{\delta_m} \right],$$

the matrix $\tilde{K}$ can be expressed by the matrix product,

$$\tilde{K} = \tilde{Y}\tilde{Y}^{\mathrm{T}}.$$

Instead of solving the $N \times N$ eigenvalue problem for $\tilde{K}$, Sirovich proposed to solve the $nm \times nm$ eigenvalue problem

$$\tilde{Y}^{\mathrm{T}}\tilde{Y}v^i = \lambda_i v^i, \qquad v^i \in \mathbb{R}^{nm}. \tag{6.11}$$

The eigenvalues of the matrix $\tilde{Y}^{\mathrm{T}}\tilde{Y}$ coincide with those of the matrix $\tilde{K}$, since multiplication of Eq. (6.11) with $\tilde{Y}$ from the left yields

$$\left(\tilde{Y}\tilde{Y}^{\mathrm{T}}\right)\tilde{Y}v^i = \lambda_i\tilde{Y}v^i.$$

Furthermore, the eigenvectors of the matrix $\tilde{K}$ can be expressed in terms of the eigenvectors of $\tilde{Y}^{\mathrm{T}}\tilde{Y}$,

$$u^i = \frac{1}{\sqrt{\lambda_i}}\tilde{Y}v^i, \tag{6.12}$$

where the scaling factor $\frac{1}{\sqrt{\lambda_i}}$ needs to be introduced due to normalization. If nm is smaller than N, the eigenvalue problem Eq. (6.11) can be solved more efficiently than the original one.

In our case, the evaluation of the original model yields time-independent steady-state solution vectors instead of time-dependent trajectories that is

$$\mathcal{Y} = \{y^1, \dots, y^n\} \subset \mathbb{R}^N.$$

The problem of finding an approximating k-dimensional subspace of $\mathbb{R}^N$ which minimizes the least-squares error in Eq. (6.2) is strongly related to the SVD of a rectangular matrix, cf. Section 2.4. To see this, let the so-called *snapshot matrix* be defined by $Y := [y^1, \dots, y^n] \in \mathbb{R}^{N\times n}$, let $r := \operatorname{rank}(Y)$ and let $Y = U\Sigma V^{\mathrm{T}}$ be a SVD of Y. The matrix K in Eq. (6.3) simplifies to

$$K = \sum_{i=1}^{n}(y^i)(y^i)^{\mathrm{T}} = YY^{\mathrm{T}}. \tag{6.13}$$

By using the SVD of Y, the matrix K can be written as,

$$K = YY^{\mathrm{T}} = U\Sigma V^{\mathrm{T}}V\Sigma^{\mathrm{T}}U^{\mathrm{T}} = U\Sigma\Sigma^{\mathrm{T}}U^{\mathrm{T}},$$

where $\Sigma\Sigma^{\mathrm{T}} = \operatorname{diag}(\sigma_1^2, \dots, \sigma_r^2, 0, \dots, 0) \in \mathbb{R}^{n\times n}$ is a diagonal matrix which contains the squared singular values of Y in descending order on it's diagonal. Since the matrix $U = [u^1, \dots, u^n]$ is orthogonal, this is an eigendecomposition of the matrix K. Thus, the non-zero eigenvalues of K are exactly the squares of the singular values $\sigma_1, \dots, \sigma_r$ of Y and the corresponding eigenvectors of K are given by the right singular vectors $u^1, \dots, u^r$ of Y. The optimal rank-k subspace $\mathcal{U}_k \subset \mathbb{R}^N$ is therefore spanned by the first k columns $\{u^1, \dots, u^k\}$ of the matrix U.

As discussed in Section 3.2, to obtain a sufficiently accurate flow solution, the number of grid points needs to be very large, which leads to a high computational effort associated with a flow simulation. As a result, the number of grid points exceeds the number of affordable sample evaluations by some orders of magnitudes, $n \ll N$. In this situation, the computation of the matrix product YY^{T} or $Y^{\mathrm{T}}Y$ should be avoided. The SVD turned out to be numerically more stable for the computation of the POD modes, [33, p. 13].

6.2. Combining POD and variable-fidelity RSM

Suppose $\mathcal{D} \subset \mathbb{R}^d$ is the domain of input variables and we aim to approximate the high-fidelity model $y_1 \colon \mathcal{D} \to \mathbb{R}^N$ with assistance of a less accurate low-fidelity model $y_2 \colon \mathcal{D} \to \mathbb{R}^N$, which describes the same quantity.

As in the case of scalar-valued responses, the multi-dimensional quantity is sampled at few high-fidelity locations $\mathcal{X} = \{x^1, \dots, x^{n_1}\} \subset \mathcal{D}$ and some low-fidelity locations $\tilde{\mathcal{X}} = \{\tilde{x}^1, \dots, \tilde{x}^{n_2}\} \subset \mathcal{D}$. Without loss of generality, cf. Remark 6.2, we consider the corresponding high- and low-fidelity responses $\mathcal{Y} = \{y^i := y_1(x^i), i = 1, \dots, n_1\}$ and $\tilde{\mathcal{Y}} = \{\tilde{y}^i := y_2(\tilde{x}^i), i = 1, \dots, n_2\}$, the so called *snapshots*, to be centered

$$\sum_{i=1}^{n_1} y^i + \sum_{i=1}^{n_2} \tilde{y}^i = 0 \tag{6.14}$$

and stored in the snapshot matrix Y,

$$Y = \left[y^1, \dots, y^{n_1}, \tilde{y}^1, \dots, \tilde{y}^{n_2}\right] \in \mathbb{R}^{N \times n} \tag{6.15}$$

where $n := n_1 + n_2$ is the total number of snapshots (we assume $N \gg n$).

Computing the first $r = \operatorname{rank}(Y)$ POD modes by performing a SVD of the snapshot matrix $Y = U\Sigma V^{\mathrm{T}}$, yields the POD basis $\{u^1, \dots, u^r\} \subset \mathbb{R}^N$ as well as corresponding singular values $\sigma_1 \geq \dots \geq \sigma_r > 0$ and the associated right singular vectors $\{v^1, \dots, v^r\} \subset \mathbb{R}^n$, cf. Sec. 6.1. If the POD basis is not truncated, every snapshot can be represented exactly by a linear combination of the POD basis modes $u^1, \dots, u^r$. Otherwise, the same holds true for the projections Y_k of the snapshot vectors onto the POD subspace $\mathcal{U}_k = \operatorname{span}\{u^1, \dots, u^k\}$. For the sake of simplicity, the following construction of the surrogate model is done for the case that the basis is not truncated. The ideas transfer to a truncated basis analogously.

As with the known case of single-fidelity surrogate models reviewed by Braconnier et al. [9], we understand the problem of constructing a variable-fidelity surrogate model as an interpolation problem of scalar-valued basis coefficients. Every high- and low-fidelity snapshot has a unique representation in the POD basis,

$$\begin{array}{ccccccccccccccc}
y^1 & = & y_1(x^1) & = & a_1^1 & u^1 & + & a_2^1 & u^2 & + & \dots & + & a_r^1 & u^r, \\
y^2 & = & y_1(x^2) & = & a_1^2 & u^1 & + & a_2^2 & u^2 & + & \dots & + & a_r^2 & u^r, \\
 & \vdots & & & & & & & & & & & & \\
y^{n_1} & = & y_1(x^{n_1}) & = & a_1^{n_1} & u^1 & + & a_2^{n_1} & u^2 & + & \dots & + & a_r^{n_1} & u^r, \\
\tilde{y}^1 & = & y_2(\tilde{x}^1) & = & \tilde{a}_1^1 & u^1 & + & \tilde{a}_2^1 & u^2 & + & \dots & + & \tilde{a}_r^1 & u^r, \\
\tilde{y}^2 & = & y_2(\tilde{x}^2) & = & \tilde{a}_1^2 & u^1 & + & \tilde{a}_2^2 & u^2 & + & \dots & + & \tilde{a}_r^2 & u^r, \\
 & \vdots & & & & & & & & & & & & \\
\tilde{y}^{n_2} & = & y_2(\tilde{x}^{n_2}) & = & \tilde{a}_1^{n_2} & u^1 & + & \tilde{a}_2^{n_2} & u^2 & + & \dots & + & \tilde{a}_r^{n_2} & u^r,
\end{array} \tag{6.16}$$

with

$$\begin{aligned}
a_i^j &= \left\langle y^j, u^i \right\rangle = \sigma_i v_j^i \quad , j = 1, \dots, n_1, i = 1, \dots, r, \\
\tilde{a}_i^j &= \left\langle \tilde{y}^j, u^i \right\rangle = \sigma_i v_{n_1+j}^i, j = 1, \dots, n_2, i = 1, \dots, r.
\end{aligned} \tag{6.17}$$

In other words, there is one i-th basis coefficient for every high- and low-fidelity sample location corresponding to the centered snapshots. These values, $a_i^1, \dots, a_i^{n_1}$ and $\tilde{a}_i^1, \dots, \tilde{a}_i^{n_2}$, can be considered as (scalar-valued) high- and low-fidelity samples of the basis coefficients at the sample locations $x^1, \dots, x^{n_1}$ and $\tilde{x}^1, \dots, \tilde{x}^{n_2}$. Thus, a variable-fidelity response surface modeling technique like Cokriging can be used to fit one surrogate model $\hat{a}_i(x)$ for every basis coefficient $a_i, i = 1, \dots, r$.

A prediction at an untried location $x \in \mathcal{D}$ is made by evaluating the coefficient models $\hat{a}_i$ at x and computing the linear combination

$$\hat{y}(x) = \hat{a}_1(x)u^1 + \hat{a}_2(x)u^2 + \cdots + \hat{a}_r(x)u^r \tag{6.18}$$

of the POD basis vectors.

The procedure to construct a variable-fidelity POD surrogate model is described in Algorithm 1.

Algorithm 1 Construction of a variable-fidelity multiple-output surrogate model

```
Input:  $y_1 \colon \mathcal{D} \to \mathbb{R}^N$                                  high-fidelity model
        $y_2 \colon \mathcal{D} \to \mathbb{R}^N$                                  low-fidelity model
        $\mathcal{X} = \{x^1, \dots, x^{n_1}\} \subset \mathcal{D}$                 high-fidelity sample set
        $\tilde{\mathcal{X}} = \{\tilde{x}^1, \dots, \tilde{x}^{n_2}\} \subset \mathcal{D}$   low-fidelity sample set
Output: $\hat{y} \colon \mathcal{D} \to \mathbb{R}^N$                              surrogate model

Compute high-fidelity snapshots $y^1 := y_1(x^1), \dots, y^{n_1} := y_1(x^{n_1})$.
Compute low-fidelity snapshots $\tilde{y}^1 := y_2(\tilde{x}^1), \dots, \tilde{y}^{n_1} := y_2(\tilde{x}^{n_2})$.
Collect snapshots in snapshot matrix $Y := [y^1, \dots, y^{n_1}, \tilde{y}^1, \dots, \tilde{y}^{n_2}]$.
Compute SVD of snapshot matrix $Y = U \Sigma V^{\mathrm{T}}$.
Determine $k = r$ with $\sigma_r > 0$ and $\sigma_{r+1} = 0$,
alternatively, determine $k := \min\{m \mid \mathrm{RIC}(m) > 1 - \varepsilon\}$ for a preset $\varepsilon > 0$.
for $i = 1, \dots, r$ do
    Compute the high- and low-fidelity basis coefficients,
        $a_i^1 = \sigma_i v_1^i, \dots, a_i^{n_1} = \sigma_i v_{n_1}^i \in \mathbb{R}$,
        $\tilde{a}_i^1 = \sigma_i v_{n_1+1}^i, \dots, \tilde{a}_i^{n_2} = \sigma_i v_{n_1+n_2}^i \in \mathbb{R}$.
    Use $(x^1, a_i^1), \dots, (x^{n_1}, a_i^{n_1})$ and $(\tilde{x}^1, \tilde{a}_i^1), \dots, (\tilde{x}^{n_2}, \tilde{a}_i^{n_2})$ to fit a variable-fidelity
    RSM $\hat{a}_i(x)$ for the $i$-th basis coefficient, e.g. via Cokriging.
end for
return $\hat{y} \colon \mathcal{D} \to \mathbb{R}^N, x \mapsto \hat{a}_1(x) u^1 + \dots + \hat{a}_k(x) u^k$.
```

6.3. Some theoretical remarks on the variable-fidelity approach

For single-fidelity Kriging interpolation, it was shown in [80, Observation 2, p. B517] that entry-by-entry interpolation of the snapshot vector entries yields exactly the same surrogate model as the interpolation of the corresponding basis coefficients—provided the same correlation parameter θ and the same regression model $f(x)$ is chosen for all Kriging models. This fundamental result applies to variable-fidelity Kriging models as well:

Lemma 6.3. *If the POD basis is not truncated, then for a fixed correlation parameter θ and the same regression model $f(x)$, entry-by-entry Kriging, Cokriging or Hierarchical Kriging interpolation of the snapshot vector entries yields exactly the same surrogate model as the interpolation of the corresponding basis coefficients.*

Proof. The result was proven in [80, Observation 2, p. B517] for the case of RBF interpolation and it was stated that the same argumentation holds for Kriging interpola-

tion. The arguments transfer to Cokriging as well:
The most important point to notice is that the Cokriging weights vector $\tilde{\lambda}(x)$ which is defined by

$$\begin{bmatrix} \tilde{\lambda}(x) \\ \tilde{\mu}(x) \end{bmatrix} = \begin{bmatrix} R & F \\ F^{\mathrm{T}} & 0 \end{bmatrix}^{-1} \begin{bmatrix} r(x) \\ f_1^{\mathrm{T}}(x) \\ 0 \end{bmatrix},$$

cf. Eq. (5.37), does not depend on the vector of sampled function values. So provided that the same sample locations $\mathcal{X}$ and $\tilde{\mathcal{X}}$ are given and the same correlation kernel $\mathcal{R}_\theta(x, \hat{x})$ as well as the same regression model $f(x)$ is considered, the Cokriging weights are exactly the same, independent of the quantity which is to be interpolated. If the vector of sampled function values at the locations $\mathcal{X}$ and $\tilde{\mathcal{X}}$ is denoted by Z, the Cokriging predictor is given by the linear combination of these sampled function values with the weights,

$$\hat{z}(x) = [Z^{\mathrm{T}}, 0] \begin{bmatrix} \tilde{\lambda}(x) \\ \tilde{\mu}(x) \end{bmatrix}.$$

Let $\hat{y}_i(x)$ denote the Cokriging predictor for the i-th component of the snapshot vectors, then entry-by-entry interpolation yields

$$\hat{y}_{\text{e.-by-e.}}(x) := \begin{bmatrix} \hat{y}_1(x) \\ \vdots \\ \hat{y}_N(x) \end{bmatrix} = \begin{bmatrix} [y_1^1, \dots, y_1^n, 0] \cdot \begin{bmatrix} \tilde{\lambda}(x) \\ \tilde{\mu}(x) \end{bmatrix} \\ \vdots \\ [y_1^1, \dots, y_1^n, 0] \cdot \begin{bmatrix} \tilde{\lambda}(x) \\ \tilde{\mu}(x) \end{bmatrix} \end{bmatrix} = [Y, 0] \begin{bmatrix} \tilde{\lambda}(x) \\ \tilde{\mu}(x) \end{bmatrix}. \tag{6.19}$$

On the other hand, the snapshot matrix Y can be rewritten in terms of the POD basis representation of the snapshots,

$$Y = [y^1, \dots, y^n] = [u^1, \dots, u^r] \cdot \begin{bmatrix} a_1^1 & & a_1^n \\ \vdots & \dots & \vdots \\ a_r^1 & & a_r^n \end{bmatrix}, \tag{6.20}$$

where for simplicity $n = n_1 + n_2$ and $a_i^{n_1+j} = \tilde{a}_i^j$ for $j = 1, \dots, n_2, i = 1, \dots, r$ compared to the notation in Eq. (6.16)–(6.17).

By this, the interpolation of the POD basis coefficients yields,

$$\begin{aligned}\hat{y}_{\text{bc}}(x) &:= [u^1, \ldots, u^r] \cdot \begin{bmatrix} \hat{a}_1(x) \\ \vdots \\ \hat{a}_r(x) \end{bmatrix} = [u^1, \ldots, u^r] \cdot \begin{bmatrix} [a_1^1, \ldots, a_1^n, 0] \cdot \begin{bmatrix} \tilde{\lambda}(x) \\ \tilde{\mu}(x) \end{bmatrix} \\ \vdots \\ [a_r^1, \ldots, a_r^n, 0] \cdot \begin{bmatrix} \tilde{\lambda}(x) \\ \tilde{\mu}(x) \end{bmatrix} \end{bmatrix} \\ &= [u^1, \ldots, u^r] \cdot \begin{bmatrix} a_1^1 & \ldots & a_1^n & 0 \\ \vdots & & \vdots & \vdots \\ a_r^1 & \ldots & a_r^n & 0 \end{bmatrix} \cdot \begin{bmatrix} \tilde{\lambda}(x) \\ \tilde{\mu}(x) \end{bmatrix} = [Y, 0] \begin{bmatrix} \tilde{\lambda}(x) \\ \tilde{\mu}(x) \end{bmatrix}, \end{aligned} \tag{6.21}$$

and thus exactly the same surrogate model as entry-by-entry interpolation of the snapshot vector components, cf. Eq. (6.19).

For Hierarchical Kriging, the same argumentation holds for the Kriging interpolation of the low-fidelity snapshots. As a result, the low-fidelity surrogate model derived by entry-by-entry Kriging interpolation of the low-fidelity snapshot vector components coincides with the one derived by the Kriging interpolation of the corresponding POD basis coefficients. Thus, since the low-fidelity surrogates are the same, the arguments transfer for the Kriging interpolation of the high-fidelity snapshots as well. □

The applicability, the benefits and drawbacks of Kriging based variable-fidelity models have been widely discussed in the literature for scalar-valued outputs, see Sec. 5.2. This result shows that the arguments extend to the proposed approach of combining POD and Kriging based variable-fidelity interpolation methods. If Cokriging or Hierarchical Kriging are believed to yield good results for some interpolation problems, the same holds true for the interpolation of vector-valued quantities by following the approach proposed above. Moreover, Lemma 6.3 justifies the use of a common POD basis for the low-fidelity and high-fidelity snapshots.

The next result addresses the relationship between the singular values and the basis coefficients. At first, observe that due to the orthonormal invariance of the Euclidean norm (see e.g. [76, p. 232]), the following lemma holds.

Lemma 6.4. *Let $\mathcal{U} \subset \mathbb{R}^N$ be an Euclidean space and $\{u^1, \ldots, u^m\}$ be an orthonormal basis of $\mathcal{U}$. Let $y = a_1 u^1 + \cdots + a_m u^m$ with $a_1, \ldots, a_m \in \mathbb{R}$ be an arbitrary vector of $\mathcal{U}$. Then,*

$$\|y\|_2 = \|a\|_2, \tag{6.22}$$

with $a = (a_1, \dots a_m) \in \mathbb{R}^m$*.*

With this lemma at hand, one can prove the following lemma:

Lemma 6.5. *Let* $Y \in \mathbb{R}^{N \times n}$ *be a snapshot matrix according to Eq. (6.15) with* $\operatorname{rank}(Y) = r$ *and corresponding (not truncated) POD basis* $\{u^1, \dots, u^r\}$*. Furthermore, let* y *be a low- or high-fidelity snapshot of* Y*, i.e.* $y = Ye_j$ *for* $j \in \{1, \dots, n\}$*, with POD basis representation* $y = a_1 u^1 + \dots + a_r u^r$*. Then, it holds,*

(i) $|a_k| \leq \sigma_k$,

(ii) $\sum_{i=k}^{r} a_i^2 \leq \sigma_k^2$,

for $k = 1, \dots, r$*.*

Proof. (i) According to Eq. (6.17), the k-th basis coefficient of the snapshot y reads $a_k = \sigma_k v_j^k$, where v_j^k is the j-th component of the right singular vector v^k. The matrix $V = [v^1, \dots, v^n]$ is orthonormal, which implies $\|V\|_2 = 1$, due to the orthonormal invariance of the Euclidean norm. As a result, the absolute value of all entries of V has to be less or equal to 1, since

$$1 = \|V\|_2 = \max_{\|z\|_2 = 1} \|Vz\|_2 \geq \|Ve_l\|_2 = \sqrt{(v_1^l)^2 + \dots + (v_n^l)^2},$$

for $l = 1, \dots, n$. Thus,

$$|a_k| = |\sigma_k v_j^k| = |\sigma_k| \cdot |v_j^k| \leq \sigma_k,$$

which proves (i).

(ii) Let $Y_k = U_k U_k^{\mathrm{T}} Y, k \leq r$ be the projection of the columns of Y onto the subspace spanned by the first k left singular vectors. Due to the Eckart-Young Theorem it holds,

$$\begin{aligned}
\sigma_{k+1}^2 &= \|Y - Y_k\|_2^2 = \max_{\|u\|_2=1} \|(Y - Y_k)u\|_2^2 \\
&= \max_{\|u\|_2=1} \|(Y - U_k U_k^{\mathrm{T}} Y)u\|_2^2 = \max_{\|u\|_2=1} u^{\mathrm{T}}(Y - U_k U_k^{\mathrm{T}} Y)^{\mathrm{T}}(Y - U_k U_k^{\mathrm{T}} Y)u \\
&= \max_{\|u\|_2=1} \left(u^{\mathrm{T}} Y^{\mathrm{T}} Y u - u^{\mathrm{T}} Y^{\mathrm{T}} U_k U_k^{\mathrm{T}} Y u - u^{\mathrm{T}} Y^{\mathrm{T}} U_k U_k^{\mathrm{T}} Y u + u^{\mathrm{T}} Y^{\mathrm{T}} U_k \underbrace{U_k^{\mathrm{T}} U_k}_{I_k} U_k^{\mathrm{T}} Y u\right) \\
&= \max_{\|u\|_2=1} \left(u^{\mathrm{T}} Y^{\mathrm{T}} Y u - u^{\mathrm{T}} Y^{\mathrm{T}} U_k U_k^{\mathrm{T}} Y u\right) \\
&\geq e_i^{\mathrm{T}} Y^{\mathrm{T}} Y e_i - e_i^{\mathrm{T}} Y^{\mathrm{T}} U_k U_k^{\mathrm{T}} Y e_i = y^{\mathrm{T}} y - y^{\mathrm{T}} U_k U_k^{\mathrm{T}} y \\
&= \|y\|_2^2 - \|U_k^{\mathrm{T}} y\|_2^2.
\end{aligned}$$

Due to Lemma 6.4, we find $\|y\|_2^2 = \|a\|_2^2 = \sum_{i=1}^r a_i^2$. Because the columns of U are orthonormal and $y \in \operatorname{span}\{u^1, \ldots, u^r\}$, the coefficients a_i of the corresponding basis representation $y = a_1 u^1 + \cdots + a_r u^r$ are determined by the inner products

$$a_i = \langle y, u^i \rangle, \qquad i = 1, \ldots, r.$$

Therefore,

$$\|U_k^{\mathrm{T}} y\|_2^2 = \|(a_1, \ldots, a_k)\|_2^2 = \sum_{i=1}^k a_i^2.$$

Overall,

$$\sigma_{k+1}^2 \geq \|y\|_2^2 - \|U_k^{\mathrm{T}} y\|_2^2 = \sum_{i=k+1}^r a_i^2$$

for $k = 1, \ldots, r-1$. In addition, it holds,

$$\sum_{i=1}^r a_i^2 = \|a\|_2^2 = \|y\|_2^2 = \|Ye_j\|_2^2 \leq \|Y\|_2^2$$

and $\|Y\|_2 = \sigma_1$, see [32, Corollary 2.4.3, p. 76]. Collectively,

$$\sum_{i=k}^r a_i^2 \leq \sigma_k^2$$

holds for $k = 1, \ldots, r$.

□

The lemma above states that the POD basis coefficients of the snapshots, and thus the sampled data for the basis coefficient models, decay with the corresponding singular values. If e.g. the singular value σ_k is close to 0, the same holds true for the k-th basis coefficient of every low- and high-fidelity sample point. As a result, each basis coefficient model operates in its own range of magnitudes bounded by the corresponding singular value and the additional error which is introduced when the basis is truncated after $k-1$ POD modes is small.

6.4. Error estimation and adaptive sampling strategies

One advantage of Kriging based response surface methods is that an estimator for the mean squared error based on the prediction of spatial correlations is provided. If such a statistical response surface model is used for the interpolation of POD basis coefficients, the following theorem holds.

Theorem 6.6. *Let $Y \in \mathbb{R}^{N\times n}$ be the snapshot matrix consisting of sample values of the high-fidelity quantity y_1 and the low-fidelity quantity y_2. Let $\{u^1,\dots,u^m\}$ be the corresponding POD basis and $U_m = [u^1,\dots,u^m]$ as before. Suppose a Kriging-related variable-fidelity response surface model, e.g. Cokriging or Hierarchical Kriging, is used to interpolate the POD basis coefficients of the snapshots to obtain a surrogate model $\hat{y}$ for the quantity y_1. The following relationship holds for the mean squared error of the surrogate model $\hat{y}$ and the mean squared error of the basis coefficient models $\hat{a}_i, i = 1,\dots,m$.*

$$\mathrm{MSE}(x) := \mathrm{E}\left[\|\hat{y}(x) - U_m U_m^{\mathrm{T}} y_1(x)\|_2^2\right] = \sum_{i=1}^{m} \mathrm{E}[(\hat{a}_i(x) - a_i)^2], \tag{6.23}$$

where $U_m U_m^{\mathrm{T}} y_1(x)$ denotes the projection of the high-fidelity functional $y_1(x)$ onto the POD subspace and a_i are the unknown true coefficient models, $i = 1,\dots,m$.

Proof. Due to the linearity of the expected value, it holds

$$\begin{aligned}
\mathrm{MSE}(x) &= \mathrm{E}\left[\|\hat{y}(x) - U_m U_m^{\mathrm{T}} y_1(x)\|_2^2\right] \\
&= \mathrm{E}\left[\|(\hat{a}^1(x) - a^1)u^1 + \dots + (\hat{a}^m(x) - a^m)u^m\|_2^2\right] \\
&= \mathrm{E}\left[\left\| U_m \cdot \begin{bmatrix} \hat{a}^1(x) - a^1 \\ \vdots \\ \hat{a}^m(x) - a^m \end{bmatrix} \right\|_2^2\right] \\
&= \mathrm{E}\left[\begin{bmatrix} \hat{a}^1(x) - a^1 & , \dots , & \hat{a}^m(x) - a^m \end{bmatrix} \underbrace{U_m^{\mathrm{T}} U_m}_{=I_m} \begin{bmatrix} \hat{a}^1(x) - a^1 \\ \vdots \\ \hat{a}^m(x) - a^m \end{bmatrix}\right] \\
&= \mathrm{E}\left[(\hat{a}^1(x) - a^1)^2 + \dots + (\hat{a}^m(x) - a^m)^2\right] \\
&= \sum_{i=1}^{m} \mathrm{E}\left[(\hat{a}^i(x) - a^i)^2\right].
\end{aligned}$$

□

This theorem allows to estimate the mean squared error of the surrogate model (6.18) based on the predicted spatial correlation of the coefficient models.

In [67], different adaptive sampling strategies for Kriging based response surface methods are compared. Two of these approaches make use of the mean squared error estimator of the Kriging method: the MSE method and the sampling sensitivity method. Due to Theorem 6.6, both methods can be directly transferred to the surrogate modeling method considered in this section. These two methods, as well as other suitable adaptive sampling strategies for the surrogate modeling approach for vector-valued outputs presented in this section are recapped in the following.

6.4.1. Leave-one-out cross-validation

The leave-one-out cross-correlation is a widely used method for model assessment, [28, Sec. 2.1.2, p. 36], [67, Sec. 4.2, p. 857]. The (high-fidelity) sample set $\mathcal{X}$ is reduced by one sample point x^i and a new surrogate model $\hat{y}_{-i}(x)$ is computed based on the remaining points $\mathcal{X} \setminus \{x^i\}$. The exact error of the surrogate model $\hat{y}_{-i}(x)$ at the neglected sample location x^i,

$$e_i = \|y_{-i}(x^i) - y^i\|, \tag{6.24}$$

can be viewed as an indicator for the information density in the neighborhood of the sample location x^i. The mean of all cross-validation errors $e_i, i = 1, \dots, n$ can be used for model assessment as a global error,

$$e = \frac{1}{n} \sum_{i=1}^{n} e_i. \tag{6.25}$$

The concept of leave-one-out cross-validation can be extended to a local error indicator by introducing the error function

$$e(x) = \frac{1}{n} \sum_{i=1}^{n} \| \hat{y}_{-i}(x) - \hat{y}(x) \|, \quad x \in \mathcal{D}. \tag{6.26}$$

A disadvantage of this error function is that it is clearly non-zero at every sample point already existing in the sample set. It should therefore be used in combination with other techniques to obtain specific locations of new sample sites. In the work of Braconnier et al. [9, § 3.4.1–3.4.2] the error function was combined with a so-called *quad-tree* data structure, to adaptively improve a Kriging-based single-fidelity surrogate model for vector-valued outputs. Following this method, new sample sites are added close to existing sites with large cross-validation error. The procedure is iterated until the cross-validation error e in Eq. (6.25) falls below a preset threshold. The authors compared their method to a priori sampling strategies like Random Latin Hypercube with the same number of sample sites and found the adaptive sampling strategy to be the most efficient among the compared.

6.4.2. The MSE method

A straight-forward idea to obtain an error measure is to use the mean squared error estimate provided by Kriging based surrogate models, [67, § 4.1, p. 857]. The MSE method selects a new sample site x^* at the location in the input parameter space $\mathcal{D}$, where the mean squared error is estimated largest:

$$x^* = \arg\max_{x \in \mathcal{D}} \mathrm{MSE}(x). \tag{6.27}$$

Due to Theorem 6.6, this sampling strategy directly transfers to the MIMO surrogate models presented above.

As also mentioned by the authors in [67], the predicted MSE is zero at each sample site and increases with growing distance to existing sample points. For that reason, the predicted MSE is suitable to detect gaps in the sampling set. Nevertheless, it should be mentioned that the prediction of the MSE only relies on the estimated spatial correlation between the given sample points. The predictor does not have any information about the real difference between the Kriging model and the original model at an untried point. Therefore, the quantitative values of the predicted MSE have to be treated with caution.

6.4.3. Sampling sensitivity method

To address the issues associated with the cross-validation error and the Kriging MSE estimator, Laurenceau et. al combined both strategies to define an adaptive sampling criterion for response surface Kriging models, [48, Sec. 2.2.2].

A new sample point is selected at

$$x^* = \arg\max_{x \in \mathcal{D}} e(x) \, \mathrm{MSE}(x), \tag{6.28}$$

where $e(x)$ is the cross-validation error defined in Eq. (6.26) and $\mathrm{MSE}(x)$ is the mean squared error estimate of the surrogate model. As with the MSE method, this idea directly transfers to the surrogate modeling technique for vector-valued quantities presented above.

The authors called their selecting criterion *sampling sensitivity method* and compared this approach to other sampling strategies. Among different approaches presented in [48], the sampling sensitivity method yields the best surrogate models for the chosen aerodynamic test case.

6.4.4. Adaptive gridding method

In [67], the sampling sensitivity method was also compared to the adaptive gridding method, introduced in [11, Sec. 5, p. 57 et seq.]. The idea of adaptive gridding, is to divide the parameter space $\mathcal{D}$ into smaller subdomains according to the estimated correlation length of the different parameters. Leave-one-out cross-validation is then performed locally at these subdomains. Whenever the cross-validation error in a subdomain exceeds a preset threshold, a new sample point is added in this subdomain. In the original article, [11, Sec. 6, p. 60 et sqq.], the new sample point is added at a location x^* which maximizes the determinant of the Kriging correlation matrix corresponding to the sample set $\mathcal{X}$ supplemented by the new point x^*. This local optimization step does not directly transfer to the multi-dimensional case described above, due to the fact that there is one correlation matrix for each basis coefficient model. Instead, for example the MSE method can be used to determine a new sample location. The adaptive gridding strategy has the advantage that it can be run completely in parallel. While the other presented methods search for global maxima of the corresponding error term and add only one new sample location per iteration, the adaptive gridding method searches for local optima of the error term in smaller subdomains.

7. Numerical experiments

In this section, the proposed techniques from Sec. 6 are investigated by means of numerical experiments. After a proof of concept for the variable-fidelity surrogate modeling of vector-valued quantities and the extended adaptive sampling MSE method in Sec. 7.1.1 and Sec. 7.1.2, we focus on the applicability of the variable-fidelity surrogate modeling technique in industrial car shape optimization. To this end, two different case studies were set up regarding a high-resolution model of a Volkswagen Passat B6 in a windtunnel[11]: For the first case study in Sec. 7.2, the rear roof of the car was parameterized via free-form deformation with five design-parameters. Low- and high-fidelity data was generated by RANS simulations. For the second case study considered in Sec. 7.3, three design parameters were introduced at the rear of the car. While RANS simulations serve as low-fidelity data generators, DES was used to produce high-fidelity data.

7.1. Proof of concept

7.1.1. Variable-fidelity surrogate modeling for vector-valued outputs

To show the general applicability of the proposed variable-fidelity surrogate modeling technique for vector-valued outputs, we consider a two-dimensional airfoil profile as a benchmark problem. The geometry and the computational grid of the airfoil were taken from the tutorial `airfoil2D` for the incompressible steady-state Navier-Stokes flow solver `simpleFoam` of the OpenFOAM project [77] (version 2.3). The computational grid corresponds to the spatial discretization of the underlying Navier-Stokes PDE and features $21,821$ points and $10,720$ cells. The grid is displayed in Figure 11.

This test case has been considered in our article [6] as well. However, the purpose in [6] was to demonstrate theoretical findings on the Cokriging response surface method. The quantity of interest in this study was the scalar-valued lift-to-drag ratio as a function of the angle of attack of the incoming flow. In the study at hand, we aim to approximate the whole pressure field in the discretized space depending on the angle of attack α.

The velocity of the incoming flow is fixed at $75\,\mathrm{m\,s^{-1}}$, while the angle of attack varies from -10° to 20°. The OpenFOAM flow solver `simpleFoam` with the Spalart-

---

[11] The computer model as well as the data were kindly provided by Volkswagen.

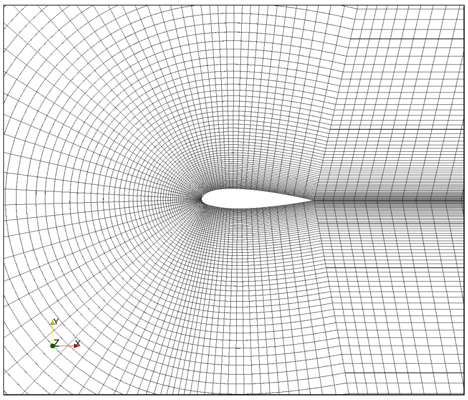

Figure 11.: Computational grid of the airfoil profile; 21,812 points

Allmaras turbulence model was used to obtain high- and low-fidelity solutions for the pressure field. For the high-fidelity data 5000 solver iterations were computed to achieve a numerical convergence level of 10^{-6} as measured by the OpenFOAM pressure residual. For the low-fidelity data, the solver iterations were stopped after only 200 iterations. Stopping the iterations at such an early stage, yields OpenFOAM pressure residuals within the range $[10^{-4}, 10^{-1}]$.

A number of $n_1 = 4$ high-fidelity snapshots were taken at the sample sites

$$\mathcal{X} := \{-10, 0, 10, 20\}.$$

In addition, $n_2 = 15$ low-fidelity snapshots were calculated at equidistant steps of 2° from $\alpha = -10^\circ$ to $\alpha = 18^\circ$,

$$\tilde{\mathcal{X}} := \{-10, -8, -6, 4, 2, 0, 2, 4, 6, 8, 10, 12, 14, 16, 18\}. \tag{7.1}$$

Note that computing all 15 low-fidelity sample points adds up to less computational effort than computing a single high-fidelity snapshot. For the sake of testing, additional high-fidelity simulations were carried out at all 27 integer angles of attack between $\alpha = 10^\circ$ to $\alpha = 20^\circ$, which do not serve as high-fidelity sample sites,

$$\mathcal{X}_{\text{test}} := \{-10 + i \mid i = 0, \dots, 30\} \setminus \mathcal{X}.$$

A detail of the high-fidelity pressure fields at the 4 high-fidelity sample sites in $\mathcal{X}$ is depicted in Figure 12.

As one can see from these exemplary results, the pressure fields vary a lot in terms

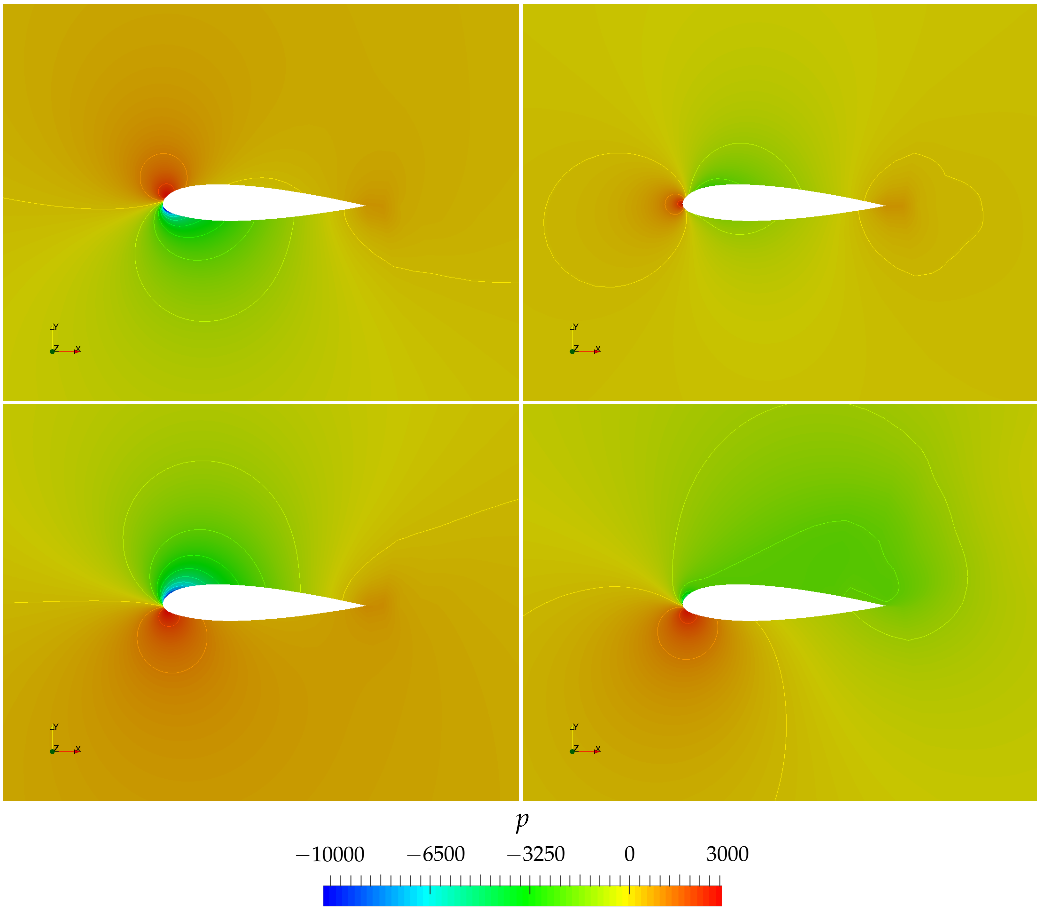

Figure 12.: High-fidelity pressure fields for the airfoil profile at the high-fidelity sample sites: angle of attack $\alpha = -10^\circ$ (upper left), $\alpha = 0^\circ$ (upper right), $\alpha = 10^\circ$ (lower left) and $\alpha = 20^\circ$ (lower right)

of stagnation points, amplitude and overall field structure, so that surrogate modeling for this single-input problem is nontrivial. For further comparison, the absolute value of the entry-by-entry difference between the high-fidelity and the low-fidelity pressure field vector for an angle of attack of $\alpha = 10^\circ$ is visualized in Figure 13.

Three surrogate models were set up following the approach presented in Section 6:

(i) A single-fidelity surrogate model considering the 4 high-fidelity sample points of the sample set $\mathcal{X}$ ("K4"). For this model, the POD basis coefficients were interpolated using Kriging with constant regression model and Gaussian correlation kernel.

(ii) A variable-fidelity surrogate model based on the 4 high-fidelity sample points of the sample set $\mathcal{X}$ and the 15 low-fidelity sample points of the sample set $\tilde{\mathcal{X}}$

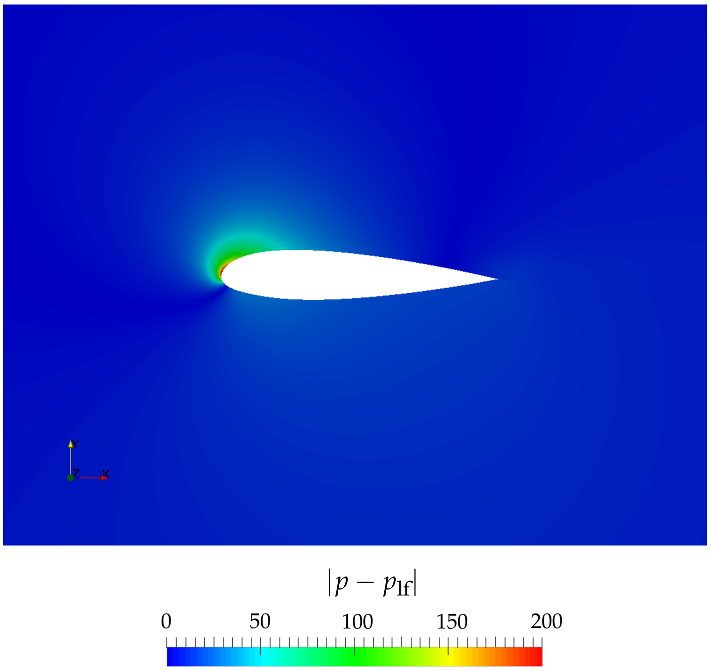

Figure 13.: Difference between the high- and low-fidelity pressure field for the angle of attack $\alpha = 10°$

("CK4+15"). For this model, Cokriging with constant regression model and Gaussian correlation kernel was used to interpolate the POD basis coefficients.

(iii) To investigate the quality of the low-fidelity sample points, a single-fidelity surrogate model was constructed considering the 15 low-fidelity sample points of the sample set $\tilde{\mathcal{X}}$. Again, a Kriging model was fit for every POD basis coefficient using a constant regression function and the Gaussian correlation kernel.

In all of the three cases, likelihood-optimal correlation parameter θ have been found by a brute force exhaustive search based on 100 sample sites per basis coefficient model[12]. For the Cokriging basis coefficient models of surrogate model CK4+15, the cross-correlation scaling factor ρ was set to $1.0 - 10^{-7}$.

All models have been evaluated at the sites of the testing set $\mathcal{X}_{\text{test}}$ and the resulting relative errors in the Euclidean norm have been computed. For the third model, which is based on the 15 low-fidelity snapshots, the averaged relative error at the

---

[12] As mentioned in Sec. 5.2.5, the numerical optimization of the correlation parameter θ is difficult as the objective function is multimodal, numerically often unstable and since the search may get stuck in local optima. In order not to deviate from the main purpose of this work and to ensure that the results are not corrupted by insufficient numerical optima, the correlation parameters θ was determined by a *brute force exhaustive search.*

testing sites sums up to 503 %, showing that the low-fidelity snapshots alone are not suitable to approximate the high-fidelity pressure fields. The relative errors of the other two surrogate models, the single-fidelity model K4 and the variable-fidelity model CK4+15, are displayed as functions of the angle of attack α in Figure 14.

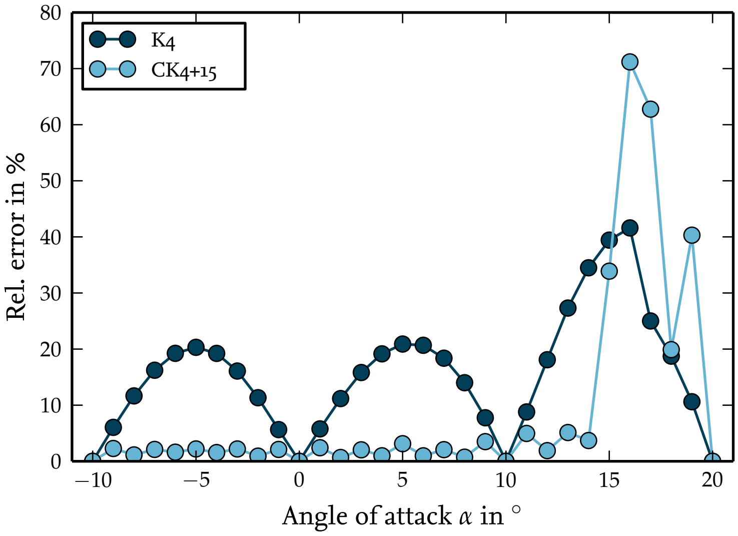

Figure 14.: Relative error of the single- and variable-fidelity surrogate model for the pressure fields of the wing profile as a function of the angle of attack α

Both models yield bad results at a high angle of attack, which can be explained by physical reasons: For a high angle of attack there are large regions of flow separation at the surface of the airfoil, which goes along with a high pressure gradient in the wake of the airfoil, [29, p. 40 et seq.]. Note that such high angles of attack are outside the realistic parameter range for commercial aircrafts at cruise conditions. While between $\alpha = -10°$ and $\alpha = 10°$ the variable-fidelity model shows a nearly constant relative error of 1.98 % on average, the relative error of the single-fidelity surrogate model strongly correlates with the distance of the assessed location to the high-fidelity sample points, with a minimum error of 5.65 % in the same range—more than twice the average error of the variable-fidelity model.

The results demonstrate the potential of variable-fidelity surrogate modeling for vector-valued quantities and motivate the further investigation for industrial applications as in the next subsections.

7.1.2. The MSE method for MIMO surrogate models

As a benchmark for the extension of the MSE adaptive sampling method to multiple outputs due to Theorem 6.6, cf. Sec. 6.4.2, we consider a 2D airfoil profile again with the pressure distribution on the surface of the airfoil as quantity of interest. In order to increase the complexity in comparison to the last test case, different input parameter are chosen: The angle of attack and the velocity of the inflow is fixed, while shape parameters are introduced which control the length and the thickness of the airfoil. The morphing is performed using the commercial pre-processing software ANSA, [7]. It allows to define so-called morphing control points and boxes which enable to deform the original mesh. Thus, the computational grids of the perturbed airfoil are topologically equivalent to the original one with $26,322$ points and $12,559$ cells. The computational grid of the unperturbed geometry and the morphing control points are displayed in Figure 15.

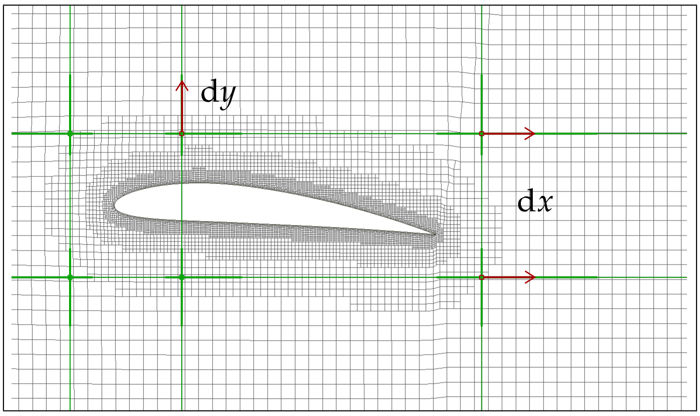

Figure 15.: Closeup of the computational grid of the unperturbed wing profile with control points for the shape modifications

The thickness of the airfoil is controlled by the parameter dy and the length by the parameter dx. The corresponding morphing control points were allowed to move up to 100 length units in the x and y direction, respectively, resulting in deformations of up to 10 % in the length and 80 % in the thickness with respect to the original airfoil geometry. The deformed airfoil profiles for the maximum variations of $(\mathrm{d}x, \mathrm{d}y) = (0, 100)$ and $(\mathrm{d}x, \mathrm{d}y) = (100, 0)$ are shown in Figure 16.

The velocity of the incoming flow was set to $75\,\mathrm{m\,s^{-1}}$ at an angle of attack of $5°$. A single-fidelity Kriging MIMO surrogate model for the pressure distribution is set up based on an initial sampling set of four sample sites. Until the cross-validation error at every sample point is below 1 %, the model is improved adaptively by adding one

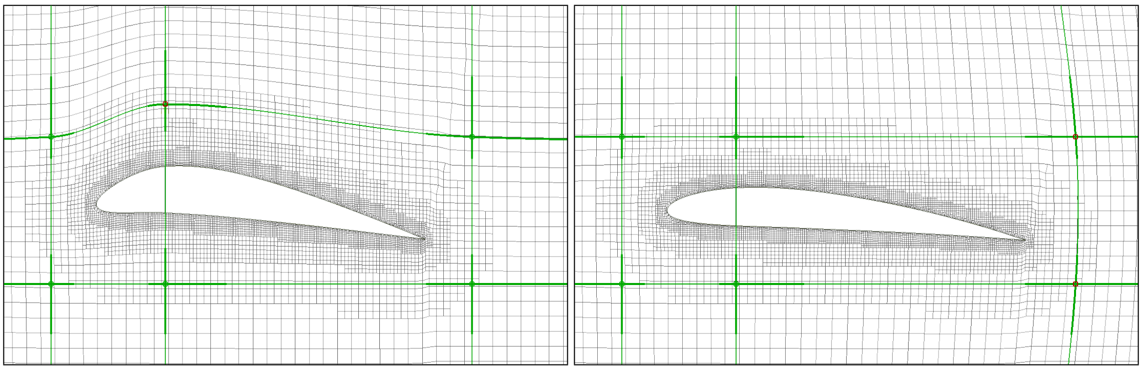

Figure 16.: Maximum shape modifications for the wing profile: $(\mathrm{d}x, \mathrm{d}y) = (0, 100)$ (left) and $(\mathrm{d}x, \mathrm{d}y) = (100, 0)$ (right)

additional sample point per iteration at the location where the MSE is estimated the highest.

As an initial sample set, a random latin hypercube sampling with 4 sample sites was set up resulting in

$$\mathcal{X}_0 = \{(100, 66.67), (66.67, 100), (33.33, 0), (0, 33.33)\}.$$

In addition, a random latin hypercube sampling with 10 sites was set up as a testing set. At each of these parameter configurations, a RANS simulation was carried out using the incompressible OpenFOAM flow solver `simpleFoam`, [77], with the Spalart-Allmaras turbulence model. To obtain the corresponding pressure distribution, the flow solver was run until convergence to an OpenFOAM residual of 10^{-6}.

An initial surrogate model for the pressure distribution on the airfoil's surface was constructed based on the responses to the sampling set $\mathcal{X}_0$. Kriging with constant regression and Gaussian correlation kernel was used to interpolate the POD basis coefficients. Likelihood-optimal values for the correlation parameter θ have been computed for every basis coefficient model via an exhaustive search on a full factorial grid of 100 sites per dimension. The mean squared error estimate of the model according to Theorem 6.6 is displayed in Figure 17.

The maximum cross-validation error was found to be 13.4 % at the parameter configuration $x^1 = (100, 66.67)$. Because this error is larger than the threshold of 1 %, a new sample point was computed at the location of the global optimum of the MSE estimate by using the GLP default solver of the python library `openopt`. The new sample location was found to be $x^5 := (0.0, 100.0)$ and a RANS simulation was performed

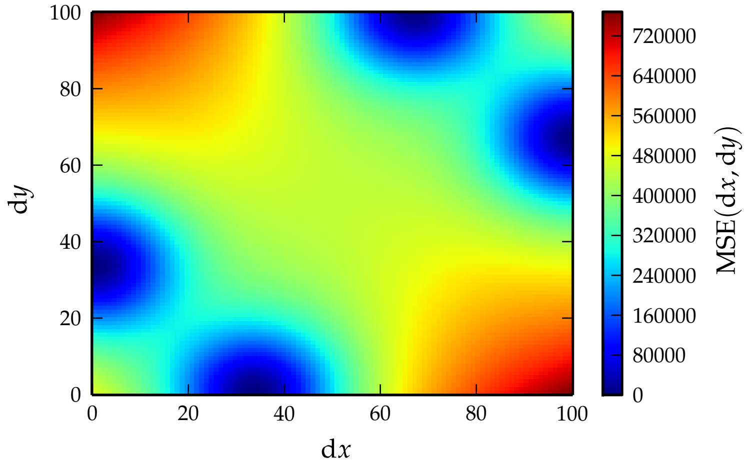

Figure 17.: Kriging-MSE estimate for the initial model K4

to obtain the pressure distribution to the corresponding morphed airfoil profile. A new surrogate was built and the procedure was iterated until after 3 iteration steps, the cross-validation error fell below 1 %. The final sample set is

$$\mathcal{X}_3 = \mathcal{X}_0 \cup \{(0.0, 100.0), (83.94, 15.60), (48.14, 55.32)\}.$$

The initial surrogate model, as well as the surrogate models of iteration 1 to 3 were evaluated at the testing sites and the relative error was computed in the Euclidean norm. The resulting average, maximum and minimum relative errors are shown in Figure 18.

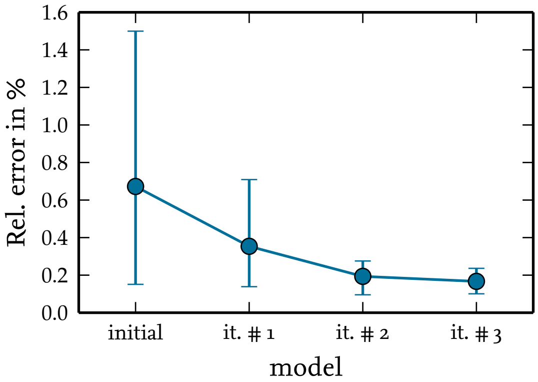

Figure 18.: Average, maximum and minimum relative error of the initial surrogate model and the models of iteration 1 to 3 with respect to the testing data set

The average relative error of initially 0.67 % for the surrogate model based on the sampling set $\mathcal{X}_0$ could be reduced to an average relative error of 0.17 % after 3 iteration steps by adding 3 additional snapshots. The range between maximum and minimum relative errors has also been reduced: From 1.34 percentage points for the initial surrogate model to 0.14 percentage points for the model after 3 iterations. Note also that the relative errors of the final surrogate model at every testing sample site is far below the preset cross-validation threshold of 1 %.

For further comparison, the entry-by-entry relative errors of the pressure distribution vectors for the 4 surrogate models at the exemplary testing configuration $(33.33, 77.78)$ are visualized in Figure 19.

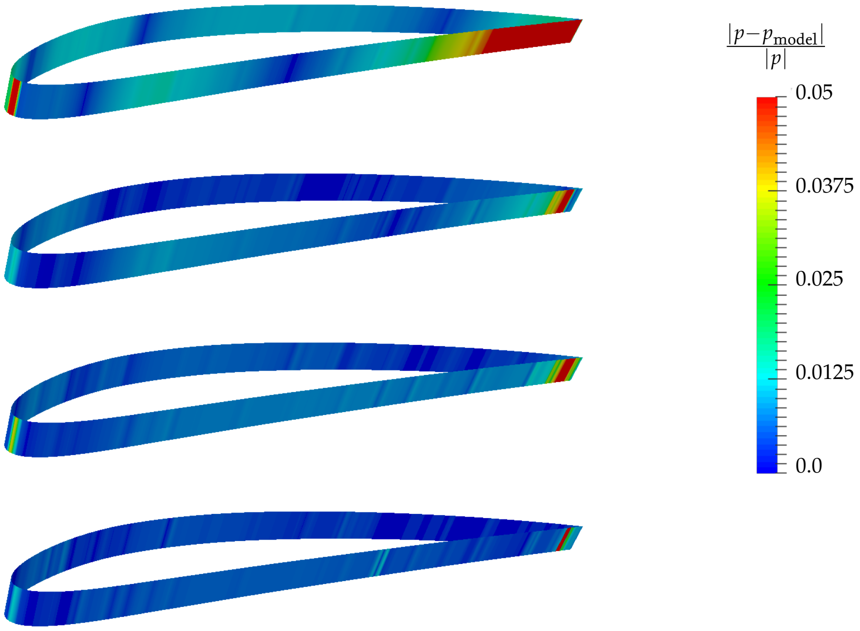

Figure 19.: Visualization of the entry-by-entry relative error of the 4 surrogate models for the testing configuration $(33.33, 77.78)$; from top to bottom: the initial model, the model after it. # 1, the model after it. # 2 and the model after it. # 3

As one can see, the regions where the entry-by-entry relative error is close to 5 %, i.e. the red colored regions at the wing profile, get smaller from iteration to iteration, while the regions with a relative error close to 0, i.e. the regions which are dark blue colored, grow. This shows that the MSE method for MIMO surrogate models is generally capable of detecting gaps in the sampling set and is able to improve the quality of the surrogate model.

In order to compare the adaptive sampling approach with a priori latin hypercube samplings, 5 additional Kriging based surrogate models were set up based on random latin hypercube samplings with 7 sample points—the same number of snapshots as for the final adaptive sampling surrogate model. The resulting average, maximum and minimum relative errors of these surrogate models ("LH1"–"LH5") as well as the final adaptive sampling surrogate model "adapt." are displayed in Figure 20.

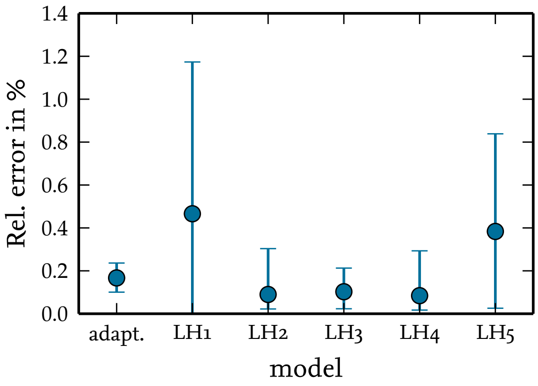

Figure 20.: Average, maximum and minimum relative errors with respect to the testing data set of the final adaptive sampling surrogate model ("adapt.") and the 5 models based on random latin hypercube samplings with the same number of snapshots ("LH1"–"LH5")

While the surrogate models LH2, LH3 and LH4 feature average relative errors which are below the average relative error of the adaptive sampling based model (0.09 %, 0.10 % and 0.08 % compared to 0.17 %), the average relative errors of model LH1 and LH5 (0.47 % and 0.38 %) are about twice the average relative error of this model. All 5 latin hypercube based models have a higher range between maximum and minimum relative error than the adaptive sampling based model: The range varies between 1.17 percentage points for model LH1 and 0.19 percentage points for model LH3, compared to a value of 0.14 percentage points for the adaptive sampling based model.

The results show that there is a huge difference in the quality of the different random latin hypercube based models. Three of the five constructed models turned out to feature a lower average relative error with respect to the testing set than the adaptive sampling based model. The adaptive sampling based model, on the other hand, features the smallest range between maximum and minimum relative error by means of the testing set. The biggest advantage of adaptive sampling strategies over a priori samplings like random latin hypercube designs is that they enable to construct surrogate models which fulfill a preset quality criterion and thus allow to control the quality of the model. By means of the underlying test case it was shown that the extension of the MSE method due to Theorem 6.6 is a suitable adaptive sampling strategy for MIMO surrogate models.

7.2. Case study 1: RANS

For the first industrial-scaled test case, a computer model of a Volkswagen Passat B6 was discretized in a wind tunnel with 17 million cells and 1 million surface faces. This test case was also considered in [61, III.B] for the *Active Subspace Method*, which aims to reduce the number of design variables to decrease the computational effort of constructing a response surface model. For details on this method the reader is referred to this paper and the references therein.

A section of the grid is displayed in Fig. 21.

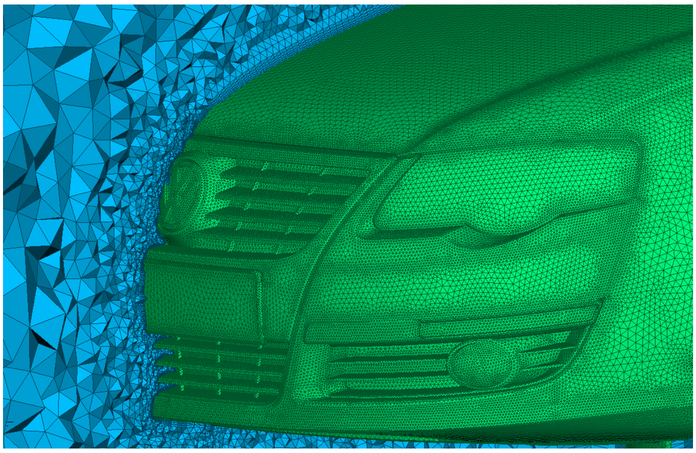

Figure 21.: Closeup of a section of the grid in case study 1

Two control points, resulting in a total of 5 design parameters, were introduced to allow free-form deformations at the rear roof of the car: The control point A at the outer edge of the roof was allowed to move in all three spatial directions, resulting in the parameters A_x, A_y and A_z. The changes were symmetrically applied to the other side of the roof. In addition, a second control point S at the center of the roof was allowed to move in the two spatial directions of the symmetry plane of the car, introducing the design parameters S_x and S_z. The positions of the control points are shown in Fig. 22.

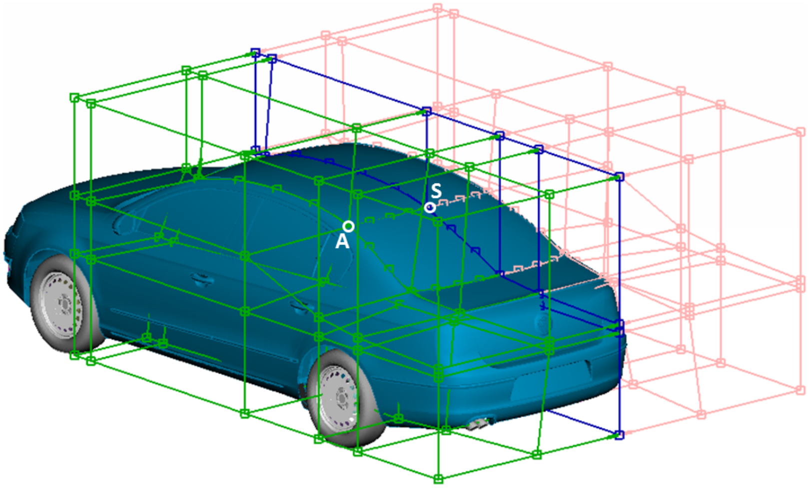

Figure 22.: The two control points at the rear roof of the car result in 5 design parameters. This Figure was also published in [61, Fig. 7b].

All 5 design parameters are varied up to a maximum of ± 2 cm compared to the original shape, as this is due to aesthetic restrictions a realistic range in car shape optimization.

As a target quantity for the surrogate models, the pressure distribution at the surface of the car was chosen. A Latin Hypercube design of experiment with 100 sample points was provided by our industrial partner. At each of these locations, a RANS simulation was carried out in OpenFOAM, [77], using the Spalart-Allmaras turbulence model. The flow solver was run for 2000 iteration steps to obtain a high-fidelity solution, which required approximately 95 core-hours of computational time per case. As low-fidelity sampled data, the partially converged flow solutions after 200 iterations of the same simulations were considered. Note that in this case low-fidelity information is available at every high-fidelity location with no additional computational effort. In order to judge the quality of the surrogate models, 10 sample locations were taken out of the Latin Hypercube sample set and were considered as reference data. Two exemplary high-fidelity pressure surface fields are shown in Fig. 23.

Three different types of surrogate models were constructed following the approach described in Sec. 6:

(i) 10 different variable-fidelity surrogate models were set up combining POD and Cokriging (“CK model # 1–10”). For each of these models, 50 sample locations were randomly chosen out of the remaining 90 Latin Hypercube sample locations. The partially converged surface fields to these 50 sample locations were treated as low-fidelity data, while the fully converged surface fields to 5 of these

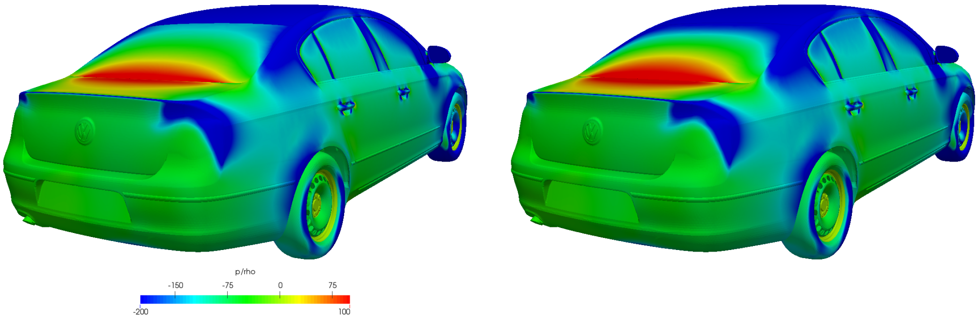

Figure 23.: Two exemplary high-fidelity snapshots for the investigated parameter space: fully converged pressure surface fields for the sample locations
$x^{15} = (0.54545, -0.50506, -1.9192, -1.87879, -2.0)$ (left),
$x^{51} = (-1.43435, -0.22223, 2.0, -0.10102, 1.75757)$ (right).
All parameter values are given in cm. For the sake of comparability of the fields, they are shown on the unperturbed geometry.

50 locations were also considered as high-fidelity data. Thus, 10 inclusive sample sets with 5 high-fidelity and 50 low-fidelity sample points were set up.

(ii) 10 variable-fidelity surrogate models were constructed combining POD and Hierarchical Kriging (“HK model # 1–10”), considering the same sample sets as the 10 surrogate models of (i).

(iii) As a reference, 10 single-fidelity surrogate models combining POD and Kriging (“K model # 1–10”) were constructed, considering 10 high-fidelity sample points respectively, since the calculation of 50 low-fidelity surface fields takes about the same computational time as the calculation of 5 additional high-fidelity surface fields. For each of these models, the 10 high-fidelity sample locations were randomly chosen from the Latin Hypercube sample set.

All surrogate models were constructed using a constant regression model and the exponential correlation kernel, cf. Tab. 1. For the Cokriging models of experiment (i), the cross-correlation scaling factor ρ was set to $1.0 - 10^{-7}$. The models were evaluated at the 10 reference locations and their average and maximum error to the RANS based solution was computed in order to compare their predictive accuracy.

The averaged relative errors in the Euclidean norm for the 10 variable-fidelity models of experiment (i) and (ii) as well as the 10 single-fidelity models of experiment (iii) with respect to the reference sampled data are displayed in Fig. 24.

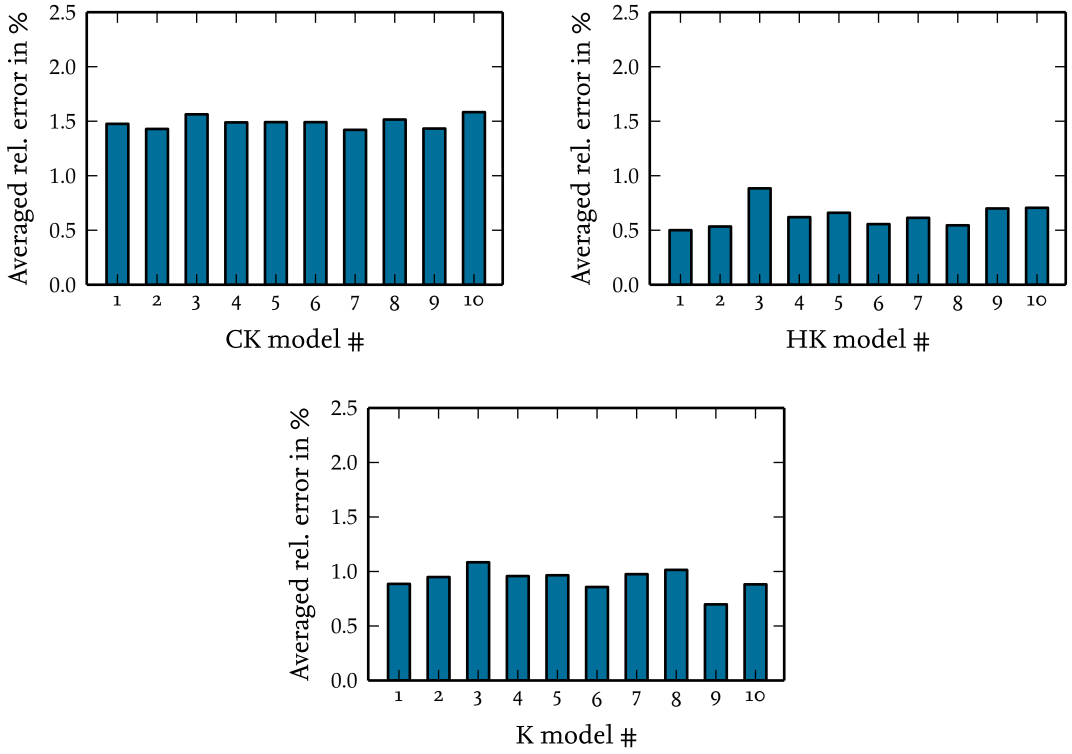

Figure 24.: Averaged relative error of CK variable-fidelity model # 1–10 (upper left), HK variable-fidelity model # 1–10 (upper right) and K single-fidelity model # 1–10 (bottom) w.r.t. the reference sampled data, respectively

The averaged relative errors of the Cokriging variable-fidelity models of experiment (i) vary between 1.41 % (CK model # 7) and 1.58 % (CK model # 10). Considering exactly the same sample sets, the Hierarchical Kriging variable-fidelity models of experiment (ii) show averaged relative errors which are of about half the size: between 0.50 % (HK model # 1) and 0.88 % (HK model # 3).

In comparison to the Kriging single-fidelity models of experiment (iii) with an error varying between 0.70 % (K model # 9) and 1.08 % (K model # 3), the predictive accuracy of the CK model is lower, while the accuracy of the HK models is considerably higher. Note that in this case the variable-fidelity surrogate models are also slightly computationally cheaper than the single-fidelity models because the low-fidelity data at the 5 high-fidelity sample points is available with no additional computational costs. This shows that the variable-fidelity surrogate modeling approach presented in this work is generally applicable to industrial-scale problems and can significantly improve the efficiency and the accuracy of surrogate models.

To further investigate the influence of the low-fidelity data on the model accuracy of the variable-fidelity surrogate models, three additional experiments were set up based on the sample set of HK model #1, which was the variable-fidelity model with the lowest averaged relative error among the models of experiment (i) and (ii):

(iv) To investigate the quality of the low-fidelity data, a single-fidelity model was constructed based on the 50 low-fidelity sample points of HK model #1 alone ("LF50").

(v) In addition, a single-fidelity surrogate model was constructed based on the 5 high-fidelity sample points of HK model #1 ("K5").

(vi) In order to study the influence of an inclusive sampling, i.e. a sampling where at some sample locations both high- and low-fidelity sampled data is available to the surrogate model, a Hierarchical Kriging variable-fidelity surrogate model was set up considering the sample set of HK model #1, but deleting the 5 double sample points from the low-fidelity data set. Thus, we end up with 5 high-fidelity and 45 low-fidelity sample points ("HK5+45").

The averaged relative error in the Euclidean norm with respect to the reference sample data of these three models and HK model #1 for comparison is shown in Figure 25.

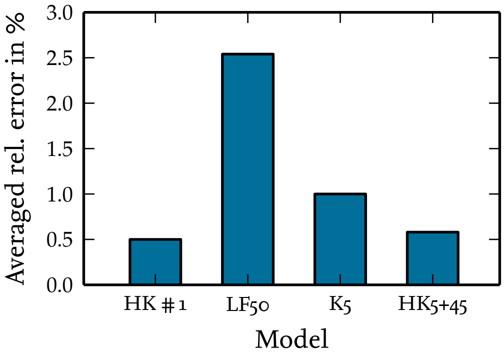

Figure 25.: Averaged relative error of the models of experiment (iv) to (vi) w.r.t. the reference sampled data, respectively

The averaged relative error of the model LF50, which is based on the low-fidelity data alone, was found to be 2.54 %, which illustrates that the quality of the low-fidelity sample data alone is not high enough to yield a comparably accurate surrogate model.

Only considering the 5 high-fidelity sample points of HK model # 1, yields an averaged relative error of 1.00 % for model K5, which is about twice as high than the error of HK model # 1. Leaving out the double sample points, one obtains a Hierarchical Kriging variable-fidelity model with an averaged relative error of 0.58 %, which is considerably smaller than for the model K5, but 16 % higher than for HK model # 1 where the 5 additional and free-of-cost low-fidelity sample points are considered.

All in all this case study demonstrates that the variable-fidelity approach presented in Sec. 6 is generally applicable to industrial scale problems and is able to improve the accuracy compared to single-fidelity surrogate models with similar computational costs regarding the offline phase. However, the single-fidelity surrogate models slightly outperform the Cokriging variable-fidelity models in this test case. Furthermore, it turned out that double sample points have a considerable impact on the accuracy of the variable-fidelity surrogate model and should therefore be included in the sample set. This is in line with the observations of Sec. 5.2.5.

7.3. Case study 2: DES combined with RANS

Closer to the operational process of investigating external aerodynamics in the Volkswagen Group is the utilization of DES-based data instead of RANS-based data, [38]. For that reason, a second case study was made to investigate whether it is possible to combine RANS and DES data in surrogate modeling. Parts of this case study have been published in our article [5]. In comparison to the case study presented therein, the numerical investigations have been extended by one additional experiment regarding Cokriging variable-fidelity models.

Because DES requires considerably finer computational grids than RANS simulations, cf. Sec. 3.2, it was not possible to use the same setting as in the previous case study. Instead in this case study, a finer grid was considered for both, high-fidelity DES and low-fidelity RANS simulations. To further increase the complexity of the test case, three design parameters were introduced at the rear end of the car instead of the rear roof: boat tailing angle, trunk length and trunk height (see Figs. 26 and 27). All three parameters were allowed to vary between $+100$ mm and -100 mm. As the edges of the free-form deformation boxes do not exactly coincide with the edges of the car, the actual deformations of the car surface are slightly below the ± 100 mm. Within this $200 \times 200 \times 200$ design space cube, several snapshot sets were created (Fig. 28):

- the reference solution in the center of the cube (Ref),
- a full-factorial data set in the eight corners of the cube (FF),
- a Latin Hypercube consisting of 20 snapshots distributed over the whole cube (LH), and
- a second Latin Hypercube with 10 snapshots, located within an interior, centered cube of 100 mm edge length (originally intended as a test data set, TT).

At all those locations in parameter space, aerodynamics computations with both RANS and DES were carried out in OpenFOAM, thereby using the Spalart-Allmaras turbulence model. The employed meshes are of the trimmed-cell type and consist of 115 million cells. Since the volume mesh is deformed along with the vehicle surface, all meshes are topologically equivalent to the one of the reference simulation. While the RANS simulations were run for 8000 iterations (requiring approximately 3,500 core-hours per case), the DES computations spanned four physical seconds (~36,000 core-hours) and were time-averaged over the last two seconds. All DES computations of the deformed geometries were, in addition, seeded with the time-averaged DES flow field of the reference configuration. Fig. 29 gives an impression of the DES flow fields themselves and the spectrum of changes within the parameter space.

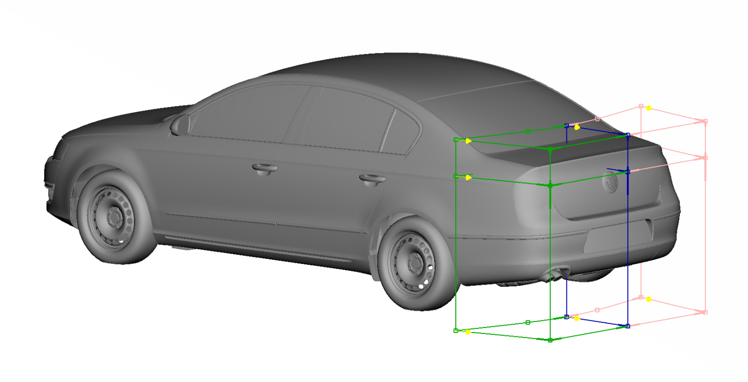

Figure 26.: Vehicle geometry and free-form deformation boxes. The green lines are the edges that control the shape deformations, and the pink edges follow their green counterparts to enforce symmetry.

The motivation for having these different data sets is to explore the trade-off between achievable accuracy of the ROM model and the associated offline computational costs. We therefore employ the data sets themselves or combinations thereof as training data for the ROM, and check its predictive capability by using either the

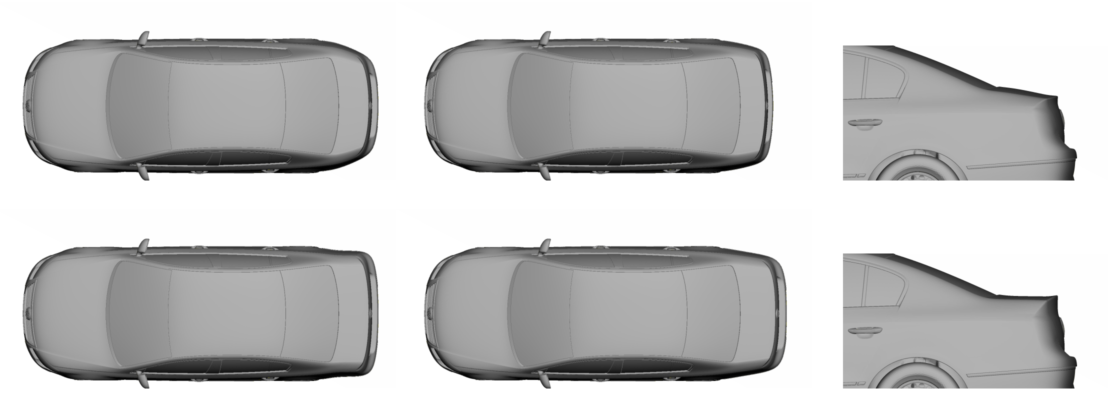

Figure 27.: Minimum (top) and maximum (bottom) shape variations for the three design variables: boat tailing angle (left), trunk length (center), and trunk height (right)

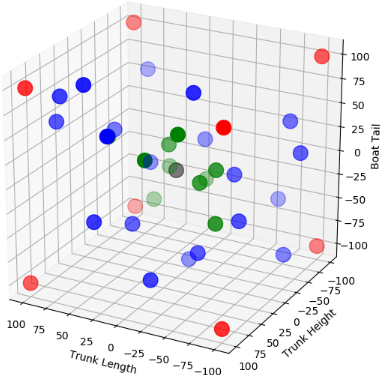

Figure 28.: The generated snapshot sets: Ref (grey, center), FF (red), LH (blue), and TT (green)

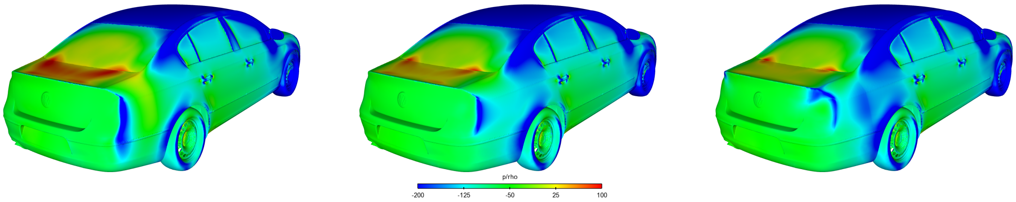

Figure 29.: Exemplary results for the investigated parameter space: time-averaged pressure for the snapshot corresponding to $(100, -100, 100)$ of the FF data set (left), the reference case (center) and the FF snapshot for $(100, 100, -100)$ (right). For the sake of comparability of the fields, they are all shown on the unperturbed geometry.

reference configuration (Ref) or the small Latin Hypercube (TT) as test data. As with case study 1, the quantity of interest is the surface field of pressure.

In the following, we apply the variable-fidelity approach described in Sec. 6 to the industrial test case introduced above. The DES-based and RANS-based flow fields serve as high-fidelity and low-fidelity data, respectively. Apart from demonstrating the general feasibility of combining DES and RANS simulation results of complex geometries into a variable-fidelity surrogate model, we are specifically interested in how any additional low-fidelity information influences the accuracy of the surrogate model.

For all experiments of this section, the surrogate models were constructed by following the respective procedures outlined in Sec. 6: While simple Kriging was used to interpolate the POD basis coefficients for the single-fidelity surrogate models, the interpolation for the variable-fidelity models was done by Hierarchical Kriging and Cokriging—in each case with constant regression and a cubic correlation kernel. The cross-correlation scaling factor ρ for the Cokriging models was set to $1.0 - 10^{-7}$. For the quantification of the predictive accuracy of each of the models, we used their average and maximum error on the 10 DES-based surface pressure fields of the TT data set.

As in case study 1, we first check whether a variable-fidelity approach can be superior to a single-fidelity model in terms of the achievable model accuracy for scenarios where the overall computational budget is fixed. To that end, the following three main experiments were set up:

(i) 10 different Hierarchical Kriging variable-fidelity surrogate models were constructed. For each of these models, we randomly chose 4 sample points out of the LH sample set and treated the corresponding DES-based surface fields as high-fidelity data, while the RANS-based data corresponding to the whole LH sample set served as low-fidelity data, so that, in total, 4 high-fidelity and 20 low-fidelity sample points were considered ("HK4+20").

(ii) 10 Cokriging variable-fidelity ROMs were constructed considering the same sample sets as the models of experiment (i), ("CK4+20").

(iii) As a DES calculation of the vehicle test case of this study takes about 10 times more computational time than a RANS simulation, the computational effort to compute the 20 low-fidelity snapshots of (i) is similar to that of 2 additional high-fidelity snapshots. For comparison with (i), we therefore constructed 10 different single-fidelity ROMs based on the DES pressure fields of 6 randomly chosen sample locations of the LH sample set ("SF6").

The averaged relative errors for the 10 variable-fidelity models of experiment (i) and (ii) as well as the 10 single-fidelity models of experiment (iii) are displayed in

Fig. 30.

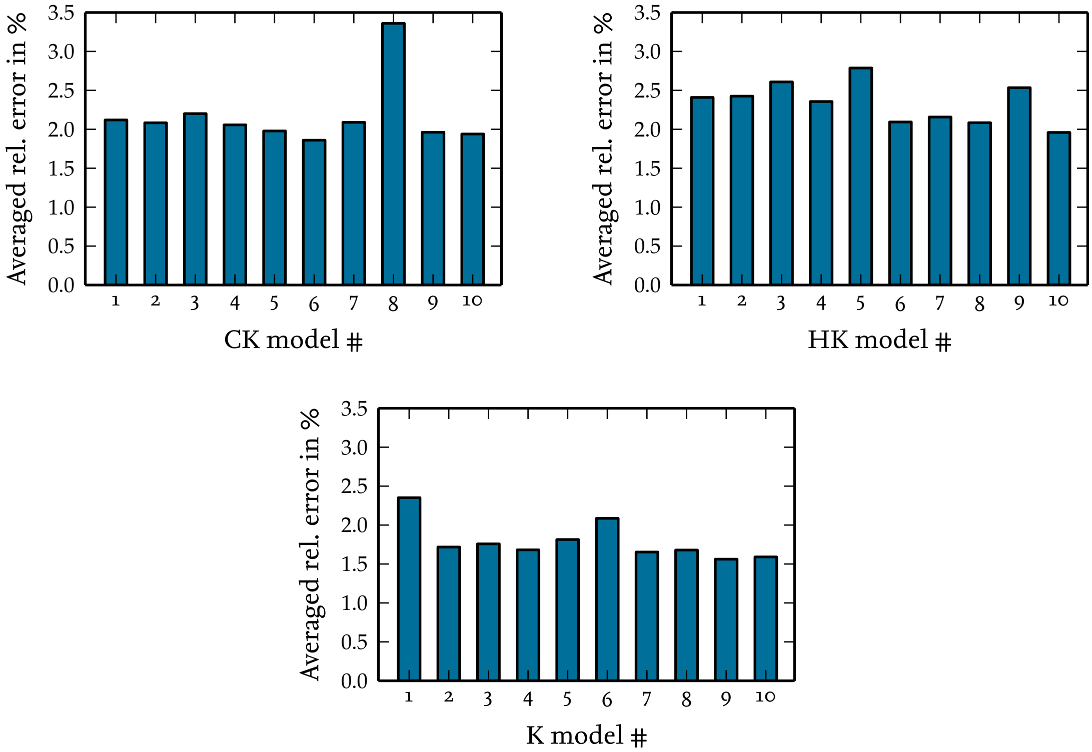

Figure 30.: Averaged relative error of CK variable-fidelity model # 1–10 (upper left), HK variable-fidelity model # 1–10 (upper right) and K single-fidelity model # 1–10 (bottom) w.r.t. the TT data set, respectively

For the Hierarchical Kriging variable-fidelity surrogate models of experiment (i), they vary between 1.96 % (model # 10) and 2.79 % (model # 5), while those for the Cokriging surrogate models of experiment (ii) are between 1.86 % (model # 6) and 3.36 % (model # 8). Apparently, when using just 4 high-fidelity snapshots, their location in parameter space has a significant impact on the quality of the ROM. The variable-fidelity model errors are also slightly higher than those obtained by the single-fidelity model of experiment (iii), which range from 1.56 % (model # 9) to 2.35 % (model # 1).

As a result of these initial experiments, we can therefore state that the variable-fidelity approach introduced in Sec. 6 is capable of combining DES and RANS snapshots into a ROM with reasonable predictive accuracy for DES test data sets of industrial complexity. For the specific case of this study, however, its accuracy was slightly outperformed by a single-fidelity ROM of comparable computational effort based entirely on DES snapshots. This behavior might be related to the respective distributions of the low-fidelity training samples LH and the high-fidelity testing samples TT: Apparently, there is no great benefit of adding low-fidelity information at the *dispersed* LH snapshots for the prediction of the rather *centered* TT samples.

In order to investigate the influence of incorporating low-fidelity information on the overall model accuracy in more detail, three more experiments were set up, cf. Sec. 7.2. As a reference for these experiments we chose the best variable-fidelity model of experiment (i), i. e. HK model # 10 (see Fig. 30), which showed the smallest averaged relative error for the TT data set among all 10 models. The training data of this reference case ("best HK4+20") was then modified as follows:

(iv) As a baseline check of the respective quality of high-fidelity vs. low-fidelity snapshots, we constructed a single-fidelity ROM on the basis of the 20 low-fidelity snapshots alone ("LF20").

(v) Another single-fidelity model ("K4") was constructed by considering only the 4 DES-based high-fidelity snapshots of the reference case.

(vi) To study the influence of double sample points—which are those sample locations where both high-fidelity as well as low-fidelity data are given—a variable-fidelity ROM was constructed where the 4 high-fidelity sample locations of the reference case were deleted from the low-fidelity sample set, so that 4 high-fidelity and 16 low-fidelity snapshots were regarded in total ("HK4+16").

As before, the models were evaluated with respect to the high-fidelity surface pressure fields at the sample locations of the TT data set. The average error thus obtained for the pure low-fidelity model LF20 of experiment (iv) is as high as 16.5 %—higher than for any other model of this study. While using only low-fidelity data was expected to be insufficient to model high-fidelity snapshots, the magnitude of this error puts the two data qualities in perspective and underpins the challenge of combining these two fidelities into a single model. It can be deduced from experiment (v) that our proposed variable-fidelity approach is indeed capable of handling such a quality contrast: Leaving out the low-fidelity data entirely from the reference case does result in a reasonable single-fidelity ROM with an average error of 2.43 % (K4). However, it is clearly inferior to the variable-fidelity reference ROM with an error of 1.96 % (best HK4+20). Thus, adding low-fidelity information to existing high-fidelity data can improve the model accuracy significantly. In this respect, experiment (vi) reveals an interesting observation on *where* to add the low-fidelity information: When removing the low-fidelity snapshots at those locations where high-fidelity is available, the error more than doubles (see Fig. 31). We therefore conclude that it is crucial for building an accurate variable-fidelity ROM to have sample locations where both high- and low-fidelity data are available. This is in line with the results of experiment (vi) in case study 1 and the theoretical findings of [6], see also Sec. 5.2.5.

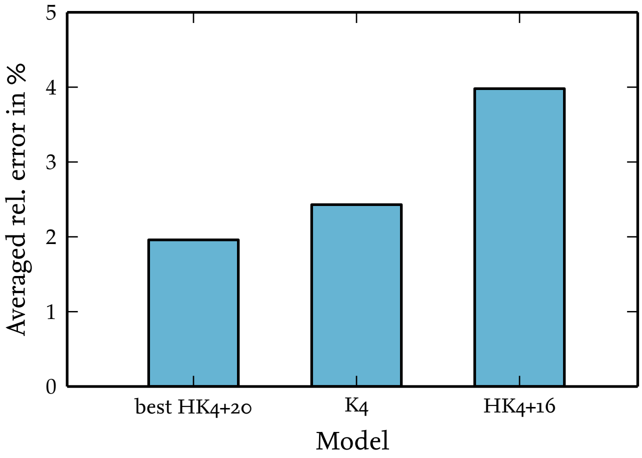

Figure 31.: Averaged relative errors w.r.t. the TT data set for the best variable-fidelity model of experiment (i) and the models of experiment (v) and (vi)

8. Conclusion and future work

In this thesis, we developed a new variable-fidelity surrogate modeling approach for the real-time prediction of vector-valued steady flow quantities. Due to their short evaluation times and the advantage of not requiring access to the source code of the flow solver, the non-intrusive data-driven approach of combining POD and a response surface interpolation technique like Kriging is of particular interest for the use in industrial vehicle shape optimization. However, the tremendous computational effort of computing high-fidelity sampled data via accurate flow simulations for realistic vehicle shapes strongly restricts the number of affordable snapshots. Thus, the applicability of established methods in this application field is limited and made the development of a new approach indispensable.

Variable-fidelity surrogate modeling is a well-studied concept in the case of scalar-valued quantities. The modeling techniques combine computational expensive high-fidelity sample values with data obtained via a second computer model which features faster evaluation times but lacks of accuracy. Although the low-fidelity data alone is not accurate enough to yield a sufficiently precise surrogate model, it can be used to obtain information about the global trend of the quantity of interest in the entire design domain. In this way it can improve the interpolation of the high-fidelity data. In the case of scalar-valued responses, variable-fidelity surrogate modeling techniques have been proven to significantly increase the efficiency of surrogate modeling. For the use in vehicle shape optimization, we extended the variable-fidelity methodology to vector-valued quantities.

As established approaches, our new variable-fidelity surrogate modeling technique relies on the idea of interpolating the scalar-valued basis coefficients of the POD basis: High- and low-fidelity snapshots are collected in a common snapshot set and a POD basis is computed based on this set. A variable-fidelity surrogate model can be fitted for every basis coefficient, e.g. by using Cokriging or Hierarchical Kriging.

It has been proven that, if the model parameters of the basis coefficient models are fixed and the POD basis is not truncated, this procedure yields exactly the same surrogate model as can be obtained via entry-by-entry variable-fidelity interpolation of the snapshots. This shows that the new approach is indeed a generalization of the well-studied variable-fidelity modeling approach for scalar-valued quantities.

Furthermore it has been proven that the mean squared error estimate provided by Kriging-based response surface modeling techniques can be transferred to the surrogate modeling of vector-valued quantities. Thus, existing adaptive sampling strategies for the case of scalar-valued outputs, which make use of the mean squared error estimate, transfer to vector-valued outputs as well.

The general applicability of the new variable-fidelity surrogate modeling technique has been demonstrated by means of a benchmark problem: The pressure distribution at the surface of a 2D wing profile has been approximated as a function of the angle of attack of the incoming flow. In this test case, it has been shown that variable-fidelity modeling clearly outperforms the established single-fidelity approach at comparable computational costs.

The transferability of one of the presented adaptive sampling strategies for scalar-valued quantities to the vector-valued case has been demonstrated at a 2D wing profile as well. While the quantity of interest remained the pressure distribution at the surface, two shape parameters have been introduced as input variables instead of a varying angle of attack. On the basis of an initial sample set, the adaptive sampling strategy was used to obtain a surrogate model which fulfilled a pre-defined quality criterion. The resulting surrogate model was compared to random latin hypercube sampling based surrogate models with the same number of sample points. While the range between smallest and largest error as well as the average error strongly varies for the random sampling based models, the adaptive sampling strategy was demonstrated to reliably result in a surrogate model of good quality.

To investigate the applicability of the new surrogate modeling approach in industrial vehicle shape optimization, we carried out two case studies. Both are based on a high-resolution computer model of a Volkswagen Passat B6 provided by our industrial partner. In each case, the goal was to approximate the pressure distribution on the surface of the vehicle as a function of design parameters which act on the surface of the car.

In the first case study, the low- and high-fidelity sampled data was obtained by RANS simulations. While the pressure distributions corresponding to fully converged RANS simulations have been considered as high-fidelity data, the simulations have been aborted at an early stage of convergence to obtain the low-fidelity data. At lower computational effort for computing the snapshots, the constructed Hierarchical Kriging based variable-fidelity surrogate models outperform the Kriging based single-fidelity surrogate models with regard to the training data set.

As detached-eddy simulations are closer to the actual industrial process chain at Volkswagen, these were used to obtain high-fidelity sampled data in the second case study. Low-fidelity sampled data was computed performing RANS simulations. In this case, the single-fidelity surrogate models revealed a smaller averaged error compared to the variable-fidelity surrogate models with regard to the training data set. A possible reason for the better performance of the single-fidelity approach in this test case might be an insufficient correlation between the DES based data and the RANS data used in the variable-fidelity modeling.

In both case studies, it has been demonstrated that double sample points, which are sample locations where both low- and high-fidelity information is provided, have a beneficial impact on the variable-fidelity surrogate model. This is also in line with the results obtained for the Cokriging interpolation of scalar-valued outputs. While the traditional Cokriging approach does not allow to invoke double sample points, we derived that slight modifications enable the use of double sample points. In addition, analytical formulas for the Cokriging parameters have been derived which improve the robustness of the Cokriging method.

Further work needs to be done regarding an indicator for the quality of the low-fidelity data and their potential to improve the high-fidelity interpolation. Adaptive sampling strategies are a mandatory instrument for constructing reliable surrogate models. However, they were not part of the two industrial case studies. Due to the tremendous computational effort of computing high-fidelity data, the data sets were provided by our industrial partner. Moreover, the computation of additional snapshots was beyond our available computing capacity. Thus, an investigative study on the adaptive sampling approach needs to be postponed to future work.

After all, the new variable-fidelity modeling approach has been demonstrated to yield accurate surrogate models. Due to their near real-time evaluation times they are applicable for the use within an interactive design process in industrial vehicle shape optimization. It was shown that variable-fidelity surrogate models have the potential to significantly reduce the computational effort of computing a surrogate model.

A. Implemented and utilized software

The software which has been used in this thesis consists of four parts, which are displayed in Figure 32.

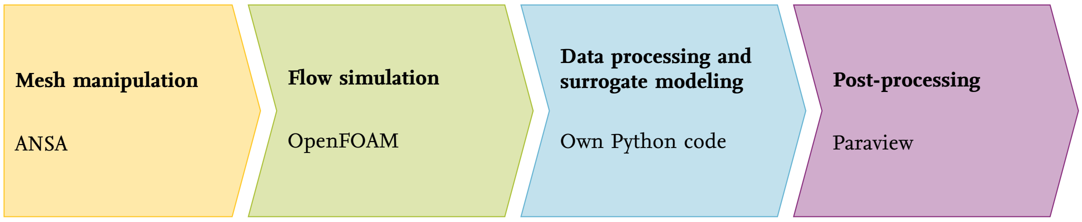

Figure 32.: Process chain and involved software

For the definition of design parameters, the morphing of the vehicle geometries and the manipulation of the computational grid, the commercial software ANSA by BETA CAE Systems, [7] has been used. ANSA is equipped with tools to perform simple design of experiment studies like random latin hypercube design or full factorial design. It is compatible with several flow solvers and automatically saves the experiments of the design study in the right format. Besides a graphical user interface, ANSA also features the possibility to perform a DOE study via the command line, which is very useful especially in the case of adaptive sampling.

The flow simulation was performed using flow solvers from the OpenFOAM library, [77]. OpenFOAM is an open source CFD toolbox with numerous compressible and incompressible flow solvers. The RANS simulation and the DES were carried out via the flow solvers `simpleFoam` and `pimpleFoam`, respectively.

The post-processing was done using the open source software Paraview, [43], which provides tools for the analysis and visualization of flow data.

For the sake of surrogate modeling, own routines have been developed in the scope of 5,000 lines of Python code. The object-oriented Python project consists of several methods and classes for the processing of data in the OpenFOAM format and the surrogate modeling of scalar- and vector-valued quantities. An overview of the different classes is given in Figure 33.

The classes can be divided in four main groups according to their function: control, data processing, surrogate modeling and miscellaneous. After a flow simulation is performed, the data is available in the OpenFOAM format. The four classes `mesh`, `openFOAMObjects`, `openFOAMOutputs` and `openFOAMProject` extract the necessary information from the OpenFOAM output files and prepare the data for the use in

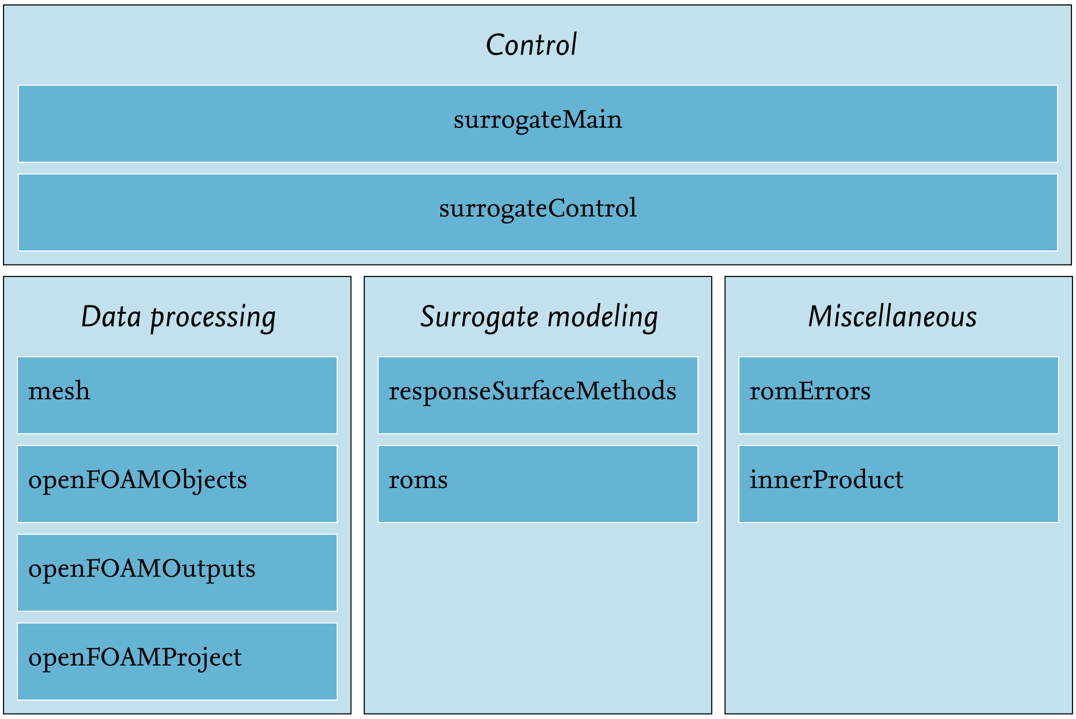

Figure 33.: Overview of the Python classes for the surrogate modeling and processing of OpenFOAM data

the surrogate modeling routines in `responseSurfaceMethods` and `roms`. The base class `responseSurfaceMethods` contains classes to fit surrogate models via Kriging, different variants of the Cokriging method, Hierarchical Kriging, Radial Basis Functions and Kriging regression. Based on these classes, the base class `roms` includes classes for the surrogate modeling of vector-valued quantities via POD combined with basis coefficient interpolation. In the case of Kriging, Hierarchical Kriging and Cokriging interpolation of the basis coefficients and methods for the adaptive sampling based on the MSE method are integrated. The methods save the computed new sample locations in an ANSA compatible format. As a result, the morphing and the flow simulation can be conveniently performed via the command line. All classes include methods to save the output of the evaluation of a surrogate model in an OpenFOAM compatible format, so that it can be processed with OpenFOAM and Paraview. Custom error messages and inner products are implemented in the classes `romErrors` and `innerProduct`, respectively. Finally, the program run is controlled via the classes `surrogateMain` and `surrogateControl`.

References

[1] A. Adam-Poupart, A. Brand, M. Fournier, M. Jerrett, and A. Smargiassi. Spatiotemporal modeling of ozone levels in Quebec (Canada): A comparison of kriging, land-use regression (LUR), and combined Bayesian maximum entropy-LUR approaches. *Environmental health perspectives*, 122(9):970–976, 2014.

[2] P. Benner, S. Gugercin, and K. Willcox. A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems. *SIAM Review*, 57(4):483–531, 2015.

[3] C. Berg, J. P. R. Christensen, and P. Ressel. *Harmonic Analysis on Semigroups*, volume 100. Springer New York, New York, NY, USA, 1984.

[4] G. Berkooz, P. Holmes, and J. L. Lumley. The Proper Orthogonal Decomposition in the Analysis of Turbulent Flows. *Annual Review of Fluid Mechanics*, 25(1):539–575, 1993.

[5] A. Bertram, C. Othmer, and R. Zimmermann. Towards Real-time Vehicle Aerodynamic Design via Multi-fidelity Data-driven Reduced Order Modeling. In *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Reston, VA, USA, 2018. American Institute of Aeronautics and Astronautics.

[6] A. Bertram and R. Zimmermann. Theoretical investigations of the new Cokriging method for variable-fidelity surrogate modeling. *Advances in Computational Mathematics*, 64(3):577, 2018.

[7] Beta CAE. ANSA pre-processor. https://www.beta-cae.com/ansa.htm. Accessed: 2018-04-04.

[8] J. Blazek. *Computational fluid dynamics: Principles and applications*. Elsevier, Amsterdam, Netherlands, 1. edition, 2001.

[9] T. Braconnier, M. Ferrier, J.-C. Jouhaud, M. Montagnac, and P. Sagaut. Towards an adaptive POD/SVD surrogate model for aeronautic design. *Computers & Fluids*, 40(1):195–209, 2011.

[10] M. D. Buhmann. *Radial basis functions: Theory and implementations*, volume 12 of *Cambridge monographs on applied and computational mathematics*. Cambridge University Press, Cambridge and New York, 2003.

[11] D. Busby, C. L. Farmer, and A. Iske. Hierarchical Nonlinear Approximation for Experimental Design and Statistical Data Fitting. *SIAM Journal on Scientific Computing*, 29(1):49–69, 2007.

[12] K. Carlberg, C. Bou-Mosleh, and C. Farhat. Efficient non-linear model reduction via a least-squares Petrov-Galerkin projection and compressive tensor approximations. *International Journal for Numerical Methods in Engineering*, 86(2):155–181, 2011.

[13] K. Carlberg, C. Farhat, J. Cortial, and D. Amsallem. The GNAT method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows. *Journal of Computational Physics*, 242:623–647, 2013.

[14] P. Chandrashekarappa and R. Duvigneau. Radial Basis Functions and Kriging Metamodels for Aerodynamic Optimization. Research Report RR-6151, INRIA, 2007.

[15] S. Chaturantabut and D. C. Sorensen. Nonlinear Model Reduction via Discrete Empirical Interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010.

[16] A. J. Chorin and J. E. Marsden. *A mathematical introduction to fluid mechanics*, volume 4 of *Texts in applied mathematics*. Springer, New York, NY, USA, 3. ed., corr. 4. printing edition, 2000.

[17] Clay Mathematics Institute. CMI Millennium Problems.
http://www.claymath.org/millennium-problems, 2018. Accessed: 2018-03-07.

[18] N. Courrier, P.-A. Boucard, and B. Soulier. Variable-fidelity modeling of structural analysis of assemblies. *Journal of Global Optimization*, 64(3):577–613, 2016.

[19] N. A. C. Cressie. The origins of kriging. *Mathematical Geology*, 22(3):239–252, 1990.

[20] N. A. C. Cressie. *Statistics for Spatial Data*. John Wiley & Sons, Inc, Hoboken, NJ, USA, 1993.

[21] C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416):953–963, 1991.

[22] L. Fahrmeir. *Multivariate statistische Verfahren.* De Gruyter, s.l., 2. edition, 1996.

[23] M. G. Fernández-Godino, C. Park, N.-H. Kim, and R. T. Haftka. Review of multi-fidelity models. http://arxiv.org/pdf/1609.07196v3.

[24] J. H. Ferziger and M. Perić. *Numerische Strömungsmechanik.* Springer Berlin Heidelberg, Berlin, Germany, 2008.

[25] A. I. J. Forrester, N. W. Bressloff, and A. J. Keane. Optimization using surrogate models and partially converged computational fluid dynamics simulations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 462(2071):2177–2204, 2006.

[26] A. I. J. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3):50–79, 2009.

[27] A. I. J. Forrester, A. Sóbester, and A. J. Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2088):3251–3269, 2007.

[28] A. I. J. Forrester, A. Sóbester, and A. J. Keane. *Engineering design via surrogate modelling: A practical guide.* Wiley, Hoboken, NJ, USA, 2008.

[29] R. W. Fox, A. T. McDonald, P. J. Pritchard, and J. C. Leylegian. *Fox and McDonald's introduction to fluid mechanics.* Wiley, Hoboken, NJ, USA, 8. edition, 2011.

[30] T. Franz, R. Zimmermann, S. Görtz, and N. Karcher. Interpolation-based reduced-order modelling for steady transonic flows via manifold learning. *International Journal of Computational Fluid Dynamics*, 28(3-4):106–121, 2014.

[31] J. Gallier. *Geometric Methods and Applications*, volume 38. Springer New York, New York, NY, USA, 2011.

[32] G. H. Golub and C. F. van Loan. *Matrix computations.* Johns Hopkins studies in mathematical sciences. Johns Hopkins Univ. Press, Baltimore, MD, USA, 4. edition, 2013.

[33] M. Gubisch and S. Volkwein. Chapter 1: Proper Orthogonal Decomposition for Linear-Quadratic Optimal Control. In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors, *Model reduction and approximation*, Computational science

and engineering, pages 3–63. Society for Industrial and Applied Mathematics, 2017.

[34] Z.-H. Han and S. Görtz. Hierarchical Kriging Model for Variable-Fidelity Surrogate Modeling. *AIAA Journal*, 50(9):1885–1896, 2012.

[35] Z.-H. Han, S. Görtz, and R. Zimmermann. Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. *Aerospace Science and Technology*, 25(1):177–189, 2013.

[36] Z.-H. Han, R. Zimmermann, and S. Görtz. A New Cokriging Method for Variable-Fidelity Surrogate Modeling of Aerodynamic Data. In *48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, Aerospace Sciences Meetings. American Institute of Aeronautics and Astronautics, 2010.

[37] Z.-H. Han, R. Zimmermann, and S. Görtz. Alternative cokriging model for variable-fidelity surrogate modeling. *AIAA Journal*, 50(5):1205–1210, 2012.

[38] M. Islam, F. Decker, E. de Villiers, A. Jackson, J. Gines, T. Grahs, A. Gitt-Gehrke, and J. Comas i Font. Application of Detached-Eddy Simulation for Automotive Aerodynamics Development. In *SAE World Congress & Exhibition*, SAE Technical Paper Series. SAE International, 2009.

[39] J.-C. Jouhaud, P. Sagaut, and B. Labeyrie. A Kriging Approach for CFD/Wind-Tunnel Data Comparison. *Journal of Fluids Engineering*, 128(4):847, 2006.

[40] A. G. Journel and C. J. Huijbregts. *Mining geostatistics*. Academic Press, London, UK, 1978.

[41] A. G. Journel and C. J. Huijbregts. *Mining geostatistics*. Blackburn, Caldwell, NJ, USA, repr. of the 5. print. 1991 edition, 2003.

[42] M. C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000.

[43] Kitware Inc. Paraview. www.paraview.org, 2018. Accessed: 2018-04-19.

[44] J. R. Koehler and A. B. Owen. Computer experiments. In S. Ghosh and C. R. Rao, editors, *Design and analysis of experiments*, volume 13 of *Handbook of Statistics*, pages 261–308. Elsevier, Amsterdam, Netherlands, 1996.

[45] U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik.* vieweg studium Aufbaukurs Mathematik. Vieweg+Teubner Verlag, Wiesbaden, Germany, 7. edition, 2003. überarbeitete Auflage.

[46] D. G. Krige. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119–139, 1951.

[47] K. Langeheinecke, P. Jany, G. Thieleke, K. Langeheinecke, and A. Kaufmann. *Thermodynamik für Ingenieure: Ein Lehr- und Arbeitsbuch für das Studium.* Springer Fachmedien Wiesbaden, Wiesbaden, Germany, 9. edition, 2013. überarb. u. erw. Aufl. 2014.

[48] J. Laurenceau and P. Sagaut. Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging. *AIAA Journal*, 46(2):498–507, 2008.

[49] L. Laurent, R. Le Riche, B. Soulier, and P.-A. Boucard. An Overview of Gradient-Enhanced Metamodels with Applications. *Archives of Computational Methods in Engineering*, 14(1):55, 2017.

[50] P. LeGresley and J. J. Alonso. Investigation of non-linear projection for POD based reduced order models for Aerodynamics. In *39th Aerospace Sciences Meeting and Exhibit*, Reston, VA, USA, 2001. American Institute of Aeronautics and Astronautics.

[51] M. A. Lifshits. *Gaussian Random Functions.* Springer Netherlands, Dordrecht, Netherlands, 1995.

[52] S. N. Lophaven, H. B. Nielsen, and J. Søndergaard. Aspects of the MATLAB Toolbox DACE: Technical Report.

[53] S. N. Lophaven, H. B. Nielsen, and J. Søndergaard. DACE: A MATLAB Kriging toolbox version 2.0: Technical Report.

[54] H. V. Ly and H. T. Tran. Modeling and control of physical processes using proper orthogonal decomposition. *Mathematical and Computer Modelling*, 33(1-3):223–236, 2001.

[55] G. Matheron. Principles of geostatistics. *Economic Geology*, 58(8):1246–1266, 1963.

[56] Michael Mifsud. *Reduced-order modelling for high-speed aerial weapon aerodynamics.* PhD Thesis, Cranfield University, Cranfield, UK, 2008.

[57] T. Mitchell, M. Morris, and D. Ylvisaker. Existence of smoothed stationary processes on an interval. *Stochastic Processes and their Applications*, 35(1):109–119, 1990.

[58] M. D. Morris, T. J. Mitchell, and D. Ylvisaker. Bayesian Design and Analysis of Computer Experiments: Use of Derivatives in Surface Prediction. *Technometrics*, 35(3):243–255, 1993.

[59] R. H. Myers, C. Anderson-Cook, and D. C. Montgomery. *Response surface methodology: Process and product optimization using designed experiments.* Wiley series in probability and statistics. John Wiley & Sons, Hoboken, NJ, USA, 4. edition, 2016.

[60] M. A. Oliver and R. Webster. Kriging: A method of interpolation for geographical information systems. *International journal of geographical information systems*, 4(3):313–332, 1990.

[61] C. Othmer, T. Lukaczyk, P. G. Constantine, and J. J. Alonso. On active subspaces in car aerodynamics. In *17th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference.* American Institute of Aeronautics and Astronautics, 2016.

[62] M. Panik. *Regression Modeling: Methods, Theory, and Computation with SAS.* Taylor and Francis, Hoboken, NJ, USA, 2009.

[63] R. Pinnau. Model Reduction via Proper Orthogonal Decomposition. In W. H. A. Schilders, H. A. van der Vorst, and J. Rommes, editors, *Model Order Reduction*, volume 13, pages 95–109. Springer Berlin Heidelberg, Berlin/Heidelberg, Germany, 2008.

[64] P. Z. G. Qian and C. F. J. Wu. Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments. *Technometrics*, 50(2):192–204, 2008.

[65] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning.* Adaptive computation and machine learning. MIT Press, Cambridge, MA, USA, 3. print edition, 2008.

[66] T. Rokita and P. P. Friedmann. Multifidelity coKriging for High-Dimensional Output Functions with Application to Hypersonic Airloads Computation. *AIAA Journal*, 56(8):3060–3070, 2018.

[67] B. Rosenbaum and V. Schulz. Comparing sampling strategies for aerodynamic Kriging surrogate models. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 92(11-12):852–868, 2012.

[68] C. W. Rowley. Model reduction for fluids, using balanced proper orthogonal decomposition. *International Journal of Bifurcation and Chaos*, 15(03):997–1013, 2005.

[69] C. W. Rowley, T. Colonius, and R. M. Murray. Model reduction for compressible flows using POD and Galerkin projection. *Physica D: Nonlinear Phenomena*, 189(1–2):115–129, 2004.

[70] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and Analysis of Computer Experiments. *Statistical Science*, 4(4):409–423, 1989.

[71] T. J. Santner, B. Williams, and W. Notz. *Design and analysis of computer experiments.* Springer series in statistics. Springer, New York, NY, USA, 2003.

[72] A. Sauerbrei and R. Zimmermann. On the Positive Definiteness of Cokriging Covariance Matrices. *PAMM*, 16(1):739–740, 2016.

[73] I. J. Schoenberg. Metric Spaces and Completely Monotone Functions. *The Annals of Mathematics*, 39(4):811, 1938.

[74] T. Schütz. *Hucho - Aerodynamik des Automobils: Strömungsmechanik, Wärmetechnik, Fahrdynamik, Komfort ; mit 49 Tabellen.* ATZ / MTZ-Fachbuch. Springer Vieweg, Wiesbaden, Germany, 6. edition, 2013. vollst. überarb. und erw. Aufl.

[75] L. Sirovich. Turbulence and the dynamics of coherent structures. I. – III. *Quarterly of Applied Mathematics*, 45(3):561–590, 1987.

[76] G. Strang. *Lineare Algebra.* Springer-Lehrbuch. Springer, Berlin/Heidelberg, Germany, 2003.

[77] The OpenFOAM Foundation. OpenFOAM: The open source CFD toolbox. www.openfoam.org, 2018. Accessed: 2018-02-22.

[78] S. P. Walton, O. Hassan, and K. Morgan. Reduced order modelling for unsteady fluid flow using proper orthogonal decomposition and radial basis functions. *Applied Mathematical Modelling*, 37(20-21):8930–8945, 2013.

[79] R. Zimmermann. Gradient-enhanced surrogate modeling based on proper orthogonal decomposition. *Journal of computational and applied mathematics*, 237(1):403–418, 2013.

[80] R. Zimmermann. A Locally Parametrized Reduced-Order Model for the Linear Frequency Domain Approach to Time-Accurate Computational Fluid Dynamics. *SIAM Journal on Scientific Computing*, 36(3):B508–B537, 2014.

[81] R. Zimmermann. An improved estimate for the condition number anomaly of univariate Gaussian correlation matrices. *Electronic Journal of Linear Algebra*, 30(1):592–598, 2015.

[82] R. Zimmermann. On the condition number anomaly of Gaussian correlation matrices. *Linear Algebra and its Applications*, 466:512–526, 2015.

[83] R. Zimmermann and S. Görtz. Non-linear reduced order models for steady aerodynamics. *Procedia Computer Science*, 1(1):165–174, 2010.

[84] R. Zimmermann and Z.-H. Han. Simplified Cross-Correlation Estimation For Multi-Fidelity Surrogate Cokriging Models. *Advances and Applications in Mathematical Sciences*, 7(2):181–202, 2010.